

COMPARISON THEORY FOR MARKOV CHAINS ON DIFFERENT STATE SPACES AND APPLICATION TO RANDOM WALK ON DERANGEMENTS

AARON SMITH

1. ABSTRACT

Let X_t and Y_t be two Markov chains, on state spaces $\Omega \subset \hat{\Omega}$. In this paper, we discuss how to prove bounds on the spectrum of X_t based on bounds on the spectrum of Y_t . This generalizes work of Diaconis, Saloff-Coste, Yuen and others on comparison of chains in the case $\Omega = \hat{\Omega}$. The main tool is the extension of functions from the smaller space to the larger, which allows comparison of the entire spectrum of the two chains. The theory is used to give quick analyses of several chains without symmetry. The main application is to a ‘random transposition’ walk on derangements.

2. INTRODUCTION

One major tool in the theory of finite state Markov chain has been the comparison technique introduced by Diaconis and Saloff-Coste in the papers [4] and [5]. This theory allows users to analyze the mixing of a Markov chain in terms of the mixing properties of another Markov chain with the same state space, as long as their stationary distributions are not too different. Practically, this may be useful because a chain of interest can be related to a similar but much simpler or more symmetric chain. In many natural examples, however, one expects Markov chains with different state spaces to have similar behaviour. For example, we might expect that removing a small number of vertices at random from a graph would generally have a small impact on the spectral gap of the associated Markov chains. The bounds in [4] and [5] do not apply in this situation. This paper is based on one way to close this gap in the literature, and we demonstrate the usefulness of this approach by deriving new bounds for several natural chains. See [10] for a useful survey on different ways to apply existing comparison techniques in different contexts. Essentially all of the techniques in that paper apply in the context of different state spaces.

Sections 3 and 4 paper deals with the theory of comparisons for distinct finite state spaces. These bounds are closely related to those found in [4] and [5]. Although the notation is quite different, they are also closely related to ideas found in the papers [8] and [9]. Those papers compared random walks on products of groups to some slightly

Date: February 11, 2013.

restricted versions, and in particular had the first special examples of comparison of Markov chains with different state spaces. To our knowledge, the only other example is Raymer's thesis [21].

Sections 5 and 6 apply the bounds obtained in the first part. We begin by looking at random walks on graphs with 'some vertices removed' that are analogous to the random walks on graphs with 'some edges removed' studied in [4]. The main application in this paper deals with a random walk on derangements, obtained by comparison with a similar random walk on permutations. This is a simple example of a Markov chain on permutations with restrictions, part of a class first studied for statistical applications in [3]. Closely related chains have been studied with very different spectral methods in [12] [1]; the same chain was studied in [15].

Sections 7-9 of this paper describe some closely related comparison bounds. This begins in section 7 with an analogous bound for discrete-time Markov chains on continuous state spaces, following the work of [24]. In section 8, we discuss the removal of some technical conditions, such as laziness. Finally, in section 9, we extend the results to another technique, the spectral profile described in [11]. We then use this to sharpen our mixing bound for an earlier example. As proved in [16], bounds obtained from the spectral profile are almost right for all chains on finite state spaces. Although we don't derive these comparisons explicitly, the discussion in section 9 applies with few changes to many other comparison inequalities based on functional analysis. An excellent survey of such bounds can be found in [20].

3. NOTATION, BACKGROUND AND STATEMENT OF RESULTS

To begin, we consider a $\frac{1}{2}$ -lazy, ergodic, irreducible, reversible Markov chain on state space $\widehat{\Omega}$, with transition kernel K and stationary distribution μ (see section 7 for remarks on obtaining related results under relaxed assumptions). We will compare this to another $\frac{1}{2}$ -lazy, ergodic, irreducible, reversible Markov chain on state space $\Omega \subset \widehat{\Omega}$, with transition kernel Q and stationary distribution ν . We will begin by comparing Dirichlet forms and Log-Sobolev constants for these two chains. Throughout, we will assume that we have satisfactory information about the chain K on the larger space, and use this to find bounds for the chain Q on the smaller space.

For a general chain on space X with kernel P and stationary distribution π , and functions f on X , we define the following functions:

$$(1) \quad V_\pi(f) = \frac{1}{2} \sum_{x,y \in X} |f(x) - f(y)|^2 \pi(x) \pi(y)$$

$$(2) \quad \mathcal{E}_P(f, f) = \frac{1}{2} \sum_{x,y \in X} |f(x) - f(y)|^2 P(x, y) \pi(x)$$

$$(3) \quad L_\pi(f) = \sum_{x \in X} |f(x)|^2 \log \left(\frac{f(x)^2}{\|f\|_{2,\pi}^2} \right) \pi(x)$$

$$(4) \quad \|f\|_{2,\pi}^2 = \sum_{x \in X} |f(x)|^2 \pi(x)$$

These quantities will be used to describe the spectral gap and log-Sobolev constants of the associated Markov chains. Recall, if P is a reversible, ergodic, irreducible, $\frac{1}{2}$ -lazy kernel, it has $|X|$ real eigenvalues satisfying

$$1 = \beta_0(P) > \beta_1(P) \geq \dots \geq \beta_{|X|-1}(P) \geq 0$$

By the variational characterization of eigenvalues, the spectral gap satisfies

$$(5) \quad 1 - \beta_1(P) = \inf_{f \neq 0} \frac{\mathcal{E}_P(f, f)}{V_\pi(f)}$$

As in [9], the log-Sobolev constant can similarly be characterized by

$$(6) \quad \alpha(P) = \inf_{f \neq 0} \frac{\mathcal{E}_P(f, f)}{L_\pi(f)}$$

Our general approach, when possible, is to use the following theorem (see Theorem 2.2 of [9]):

Theorem 1 (Mixing Time Bound via Spectral Gap and Log-Sobolev Constant). *For a $\frac{1}{2}$ -lazy reversible Markov chain X_t started at $X_0 = x$, and for $t > 1 + \frac{c}{1-\beta_1(P)} + \frac{1}{4\alpha(P)} \log \log(\frac{1}{\pi(x)})$,*

$$\|\mathcal{L}(X_t) - \pi\| \leq 2e^{-c}$$

When the log-Sobolev constant $\alpha(P)$ is available, this is often better than the usual bound in terms of just the spectral gap (see Theorem 12.3 of [18]), which gives, for $t > \frac{c}{1-\beta_1(P)} \log(\frac{1}{\pi(x)})$, the bound

$$(7) \quad \|\mathcal{L}(X_t) - \pi\| \leq 2e^{-c}$$

We will see shortly that, for many examples, it will be easy to find a very reasonable bound for $\alpha(P)$ after doing the work needed to bound $\beta_1(P)$.

It is now time to compare the functionals described in equation (1). For the remainder of this note, f will denote a function on Ω , and \widehat{f} will denote a function on $\widehat{\Omega}$ satisfying $\widehat{f}(x) = f(x)$ for all $x \in \Omega$. We call such a function an *extension* of f . We note that the inequalities

$$\begin{aligned} V_\nu(f) &\leq C_1 V_\mu(\widehat{f}) \\ L_\nu(f) &\leq C_2 L_\mu(\widehat{f}) \\ \mathcal{E}_K(\widehat{f}, \widehat{f}) &\leq C_3 \mathcal{E}_Q(f, f) \end{aligned}$$

together with the variational characterizations of β_1 and α given in equations (6) and (5) imply the following bounds on $\beta_1(Q)$ and $\alpha(Q)$ in terms of $\beta_1(K)$ and $\alpha(K)$:

$$\begin{aligned} 1 - \beta_1(Q) &\geq \frac{1}{C_1 C_3} (1 - \beta_1(K)) \\ \alpha(Q) &\geq \frac{1}{C_2 C_3} \alpha(K) \end{aligned}$$

Finding a good value for C_3 is difficult and the main object of this paper, but reasonable bounds on C_1 and C_2 can be found immediately. The following lemma is useful when μ and ν assign similar values to all points in Ω , which is the case for many natural examples.

Lemma 2 (Comparison of Variance and Log-Sobolev Constants). *Let \widehat{f} be any extension of f , and let $C = \sup_{y \in \Omega} \frac{\nu(y)}{\mu(y)}$. Then*

$$\begin{aligned} V_\nu(f) &\leq C V_\mu(\widehat{f}) \\ L_\nu(f) &\leq C L_\mu(\widehat{f}) \end{aligned}$$

Proof. Define, for c real (respectively real and strictly positive), the following functionals:

$$\begin{aligned} V_\pi(f, c) &= \sum_{x \in X} |f(x) - c|^2 \pi(x) \\ L_\pi(f, c) &= \sum_{x \in X} (|f(x)|^2 \log(|f(x)|)^2 - |f(x)|^2 \log(c) - |f(x)|^2 + c) \pi(x) \end{aligned}$$

Recall that $V_\pi(f) = \inf_{c \in \mathbb{R}} V_\pi(f, c)$, and it is shown in [13] that $L_\pi(f) = \inf_{c \in \mathbb{R}, c > 0} L_\pi(f, c)$. Thus, we can write

$$\begin{aligned}
V_\nu(f, c) &= \sum_{x \in \Omega} |f(x) - c|^2 \nu(x) \\
&= \sum_{x \in \hat{\Omega}} |\hat{f}(x) - c|^2 \nu(x) \\
&= \sum_{x \in \hat{\Omega}} |\hat{f}(x) - c|^2 \frac{\nu(x)}{\mu(x)} \mu(x) \\
&\leq CV_\mu(\hat{f}, c)
\end{aligned}$$

which implies $V_\nu(f) = \inf_{c \in \mathbb{R}} V_\nu(f, c) \leq \inf_{c \in \mathbb{R}} V_\mu(\hat{f}, c) = CV_\mu(\hat{f})$. An analogous calculation shows that $L_\nu(f, c) \leq CL_\mu(\hat{f}, c)$, which implies $L_\nu(f) \leq CL_\mu(\hat{f})$. \blacksquare

As with the extension theory built up from [4], it is possible to get bounds on the entire spectrum of Q , rather than just the second-largest eigenvalue. Unlike that case, this will require the extensions to have some structure. In particular, fix a map M from \mathbb{R}^Ω to $\mathbb{R}^{\hat{\Omega}}$ so that for all $f \in \mathbb{R}^\Omega$, $Mf \in \mathbb{R}^{\hat{\Omega}}$ is an extension of f . Assume that we can show

$$(8) \quad \mathcal{E}_K(Mf, Mf) \leq C_3 \mathcal{E}_Q(f, f)$$

for all $f \in \mathbb{R}^\Omega$. Next, consider a Hermitian matrix P with real eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$, and define for any subspace W the functions

$$\begin{aligned}
L(W) &= \min \left\{ \frac{\langle Pf, f \rangle}{\langle f, f \rangle} : f \in W \right\} \\
U(W) &= \max \left\{ \frac{\langle Pf, f \rangle}{\langle f, f \rangle} : f \in W \right\}
\end{aligned}$$

Then recall from e.g. page 185 of [14] that the eigenvalues of P satisfy

$$\lambda_i = \max \{ L(W) : \dim(W^\perp) = i \} = \min \{ U(W) : \dim(W) = i + 1 \}$$

This variational characterization, together with inequality (8) and $C_1 = \sup_{y \in \Omega} \frac{\nu(y)}{\mu(y)}$, gives the bounds

$$\begin{aligned}
1 - \beta_i(Q) &\geq \frac{1}{C_1 C_3} (1 - \beta_i(K)) \\
\alpha(Q) &\geq \frac{1}{C_1 C_3} \alpha(K)
\end{aligned}$$

The main difficulty will be to bound the Dirichlet forms \mathcal{E}_Q and \mathcal{E}_K . We begin by restricting our attention to the special class of simple random walks on regular graphs, and then write a bound for general finite Markov chains.

Assume that K is a $\frac{1}{2}$ -lazy simple random walk on $\widehat{\Omega}$, with associated graph $\widehat{G} = (\widehat{V}, \widehat{E})$. That is, the kernel is given by:

$$K(x, y) = \begin{cases} \frac{1}{2} & \text{if } y = x \\ \frac{1}{2d} & \text{if } (x, y) \in \widehat{E} \\ 0 & \text{otherwise} \end{cases}$$

Then let $G = (V, E)$ be a subgraph of \widehat{G} , where V is obtained from \widehat{V} by removing m vertices, and E is obtained from \widehat{E} by removing all edges in \widehat{E} adjacent to one of the removed edges. Then let Q be a random walk on G described by

$$Q(x, y) = \begin{cases} \frac{1}{2}(2 - \frac{1}{d} \deg(x)) & \text{if } y = x \\ \frac{1}{d} & \text{if } (x, y) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $\deg(x)$ is the number of vertices in G adjacent to x . Q is the *Metropolis-Hastings* walk associated with base walk K and target distribution uniform on G (see [19] for an introduction to the Metropolis-Hastings algorithm).

To describe the comparison, it will be necessary first to choose a specific extension \widehat{f} of f . For each vertex $x \in \widehat{G}$, fix some probability measure $P_x[y]$ on G , requiring $P_x[y] = \delta_x[y]$ for $x \in G$. This defines a family of extensions by

$$(9) \quad \widehat{f}(x) = \sum_{y \in G} P_x[y] f(y)$$

Next, for each pair $(x, y) \in \widehat{E}$, fix a joint measure $P_{x,y}[a, b]$ on $G \times G$ satisfying $\sum_a P_{x,y}[a, b] = P_y[b]$ for all $b \in G$ and $\sum_b P_{x,y}[a, b] = P_x[a]$ for all $a \in G$. This is a coupling of the distributions P_x, P_y .

Next, for each $a, b \in G$ with $\sum_{x,y \in \widehat{G}} P_{x,y}[a, b] > 0$, it is necessary to define a *flow* in G from a to b . To do so, call a sequence of vertices $\gamma = \{a = v_{0,a,b}, v_{1,a,b}, \dots, v_{k[\gamma],a,b} = b\}$ a *path* from a to b if $(v_{i,a,b}, v_{i+1,a,b}) \in E$ for all $0 \leq i < k[\gamma]$. Then let $\Gamma_{a,b}$ be the collection of all paths from a to b . Call a function F from paths to $[0, 1]$ a *flow* if $\sum_{\gamma \in \Gamma_{a,b}} F[\gamma] = 1$ for all $a, b \in \Omega$. We will often write $G_{a,b}$ for the restriction of F to $\Gamma_{a,b}$. Finally, for a path $\gamma \in \Gamma_{a,b}$, we will label its initial and final vertices by $i(\gamma) = a$, $o(\gamma) = b$.

For fixed measures $\{P_x\}_{x \in \widehat{G}}$, couplings $\{P_{x,y}\}_{(x,y) \in \widehat{E}}$, and flow F , we obtain the following bound on Dirichlet forms:

Theorem 3 (Comparison of Dirichlet Forms for Metropolized Simple Random Walk). *For flows, distributions, and paths as described above,*

$$\mathcal{E}_K(\hat{f}, \hat{f}) \leq \frac{n-m}{n} \mathcal{A} \mathcal{E}_Q(f, f)$$

where

$$\begin{aligned} \mathcal{A} = & \sup_{(q,r) \in E} (1 + 2 \sum_{\gamma \ni (q,r)} k[\gamma] F[\gamma] \sum_{y \notin G} P_y[o(\gamma)] \\ & + \sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] \sum_{(x,y) \in \hat{E}, x,y \notin G} P_{x,y}[i(\gamma), o(\gamma)]) \end{aligned}$$

For general Markov chains K and Q , define a graph \hat{G} with vertex set $\hat{\Omega}$ associated with K by creating an edge $(x, y) \in \hat{E}$ if $K(x, y) > 0$, and a graph G with vertex set Ω associated with Q by creating an edge $(x, y) \in E$ if $Q(x, y) > 0$. The same setup then gives the following bound:

Theorem 4 (Comparison of Dirichlet Forms for General Chains). *For flows, distributions and couplings as described above,*

$$\mathcal{E}_K(\hat{f}, \hat{f}) \leq \mathcal{A} \mathcal{E}_Q(f, f)$$

where

$$\begin{aligned} \mathcal{A} = & \sup_{Q(q,r) > 0} \frac{1}{Q(q,r)\nu(q)} \left(\sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] K(i(\gamma), o(\gamma)) \mu(i(\gamma)) \right. \\ & + 2 \sum_{\gamma \ni (q,r)} k[\gamma] F[\gamma] \sum_{y \notin G} P_y[o(\gamma)] K(i(\gamma), y) \mu(i(\gamma)) \\ & \left. + \sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] \sum_{(x,y) \in \hat{E}, x,y \notin G} P_{x,y}[i(\gamma), o(\gamma)] K(x, y) \mu(x) \right) \end{aligned}$$

We will now describe some applications of Theorem 3. The first is analogous to example 2.1 of [5]. Let K be the kernel of the $\frac{1}{2}$ -lazy simple random walk on the torus $G = \mathbb{Z}_n^2$ with edges of the form $((i, j), (i+1, j))$ and $((i, j), (i, j+1))$. Then let $v_1, v_2, \dots, v_m \in G$ be any collection of vertices with the property that no two are in the same square $\{(i, i), (i+1, i), (i, i+1), (i+1, i+1)\}$ in G . Then let Q be the Metropolis-Hastings walk associated with $G \setminus \{v_1, v_2, \dots, v_m\}$. The following is a general bound on the Dirichlet form of Q :

Theorem 5 (Comparison for Random Walk on the Torus with Holes). *All functions f on $G \setminus \{v_1, v_2, \dots, v_m\}$ have extensions \hat{f} to G so that*

$$\mathcal{E}_K(\hat{f}, \hat{f}) \leq 6 \left(1 - \frac{m}{n}\right) \mathcal{E}_Q(f, f)$$

Not all bounds are so useful. For example, if the removed vertices are of the form $\{(1, 1), (2, 2), \dots, (n-1, n-1)\} \cup \{(1, 1 + \frac{n}{2}), (2, 2 + \frac{n}{2}), \dots, (n-1, n+1 - \frac{n}{2})\}$, we have:

Theorem 6 (Comparison for Random Walk on the Torus with Bottleneck). *All functions f on $G \setminus \{(1, 1), (1, 1 + \frac{n}{2}), (2, 2 + \frac{n}{2}), \dots, (n-1, n-1)(n-1, n+1 - \frac{n}{2})\}$ have extensions \hat{f} to G so that*

$$\mathcal{E}_K(\hat{f}, \hat{f}) \leq 8n^2 \mathcal{E}_Q(f, f)$$

The result is the same upper bound as is given directly by Cheeger's inequality (see Theorem 13.14 of [18]). As discussed immediately after the proof, it seems impossible to do any better by comparison to the standard simple random walk on the torus using Theorem 3.

The main example in this paper is an application of Theorem 3 to the problem of sampling from derangements. Recall that a permutation $\sigma \in S_n$ is called a *derangement* if, for all $i \in [n]$, $\sigma(i) \neq i$. We will compare the well-known 'random transposition' walk on S_n to its restriction to the derangements D_n . More precisely, consider the Cayley graph \hat{G} with vertex set $\hat{V} = S_n$ and edge set \hat{E} given by $(x, y) \in \hat{E}$ if and only if $y^{-1}x$ is a transposition. We will compare the $\frac{1}{2}$ -lazy transition kernel K on \hat{G} to its Metropolized version Q on the restriction to derangements $D_n \subset S_n$.

Although sampling from the set of derangements is not hard (it is easy to sample from S_n and rejection-sampling based on this is fairly efficient), the Markov chain is closely related to several more difficult sampling problems. There has been a great deal of interest in the problem of sampling from permutations with restrictions, beginning with the work of Diaconis, Graham and Holmes in [3]. See also the recent work [2], [1], and [15] and the references contained therein for a discussion of other examples. Our main result is:

Theorem 7 (Dirichlet Form Comparison for the Random Transposition Walk on Derangements). *Fix $n \geq 10$. All functions f on D_n have extensions \hat{f} to S_n so that*

$$\mathcal{E}_K(\hat{f}, \hat{f}) \leq 22(e+1)(1+\epsilon_n)\mathcal{E}_Q(f, f)$$

where $|\epsilon_n| \leq \frac{13}{n}$. In the other direction, any function f on D_n and any extension \hat{f} of f to S_n satisfies

$$\mathcal{E}_K(\hat{f}, \hat{f}) \geq \frac{1}{2e}\mathcal{E}_Q(f, f)$$

We will show that this easily gives the following bound on the mixing time, improving earlier bounds of $O(n^3 \log(n))$ [15]:

Corollary 8 (Mixing Properties of the Random Transposition Walk on Derangements). *The random walk described above has spectral gap satisfying*

$$1 - \beta_1(Q) = \Omega\left(\frac{1}{n}\right)$$

and log-Sobolev constant

$$\alpha(Q) = \Omega\left(\frac{1}{n \log(n)}\right)$$

By Theorem 1, there exists some constant $a > 0$ and function $f(C)$ such that $\lim_{C \rightarrow \infty} f(C) = 0$ and for $t = Cn + an \log(n)^2$, $\|\mathcal{L}(X_t) - \pi\|_{TV} \leq f(C)$.

In section 8, bounds similar to Theorem 4 are developed for discrete-time chains on continuous state spaces. The development follows the discrete theory closely, much as W. K. Yuen's development of comparison theory on continuous state spaces in [24] follows the discrete theory in [5].

Next, in section 9, we briefly discuss how these extension ideas interact with a recent and powerful way of looking at Dirichlet forms, the spectral profile. The main results from [11] will be introduced. They will then be used to prove the following improvement of Theorem 5:

Theorem 9 (Improved Comparison for Random Walk on the Torus with Holes). *If X_t is a Markov chain as described in Theorem 5, we have for $t = Cn^2$,*

$$\|\mathcal{L}(X_t) - \nu\|_{TV} \leq f(C)$$

for some function f independent of n and the particular vertices removed, with $\lim_{C \rightarrow \infty} f(C) = 0$.

In particular, these random walks have a mixing time that is $O(n^2)$. This is a substantial improvement on the bound of $O(n^2 \log(n))$ obtained by a direct application of inequality (7), and a small improvement on the bound of $O(n^2 \log(\log(n)))$ found by a careful application of Theorem 1.

4. SPECTRAL GAP AND LOG-SOBOLEV ESTIMATES

In this section, we prove Theorem 4:

Proof. Assume without loss of generality that no paths contain repeated edges, and write

$$\begin{aligned}
\mathcal{E}_K(\widehat{f}, \widehat{f}) &= \frac{1}{2} \sum_{x,y \in \widehat{\Omega}} |\widehat{f}(x) - \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&= \frac{1}{2} \sum_{x,y \in \Omega} |f(x) - f(y)|^2 K(x, y) \mu(x) + \sum_{x \in \Omega, y \notin \Omega} |f(x) - \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&\quad + \frac{1}{2} \sum_{x,y \notin \Omega} |\widehat{f}(x) - \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&\equiv \frac{1}{2} R_1 + R_2 + \frac{1}{2} R_3
\end{aligned}$$

The goal is to compare this to $\mathcal{E}_Q(f, f) = \frac{1}{2} \sum_{x,y \in \Omega} |f(x) - f(y)|^2 Q(x, y) \nu(x)$. We begin by looking at R_1 :

$$\begin{aligned}
R_1 &= \sum_{x,y \in \Omega} |f(x) - f(y)|^2 K(x, y) \mu(x) \\
&= \sum_{x,y \in \Omega} \left| \sum_{\gamma \in \Gamma_{x,y}} F[\gamma] \sum_{i=0}^{k[\gamma]-1} (f(v_{x,y,i+1}) - f(v_{x,y,i})) \right|^2 K(x, y) \mu(x) \\
&\leq \sum_{x,y \in \Omega} \sum_{\gamma \in \Gamma_{x,y}} F[\gamma] \left| \sum_{i=0}^{k[\gamma]-1} (f(v_{x,y,i+1}) - f(v_{x,y,i})) \right|^2 K(x, y) \mu(x) \\
&\leq \sum_{x,y \in \Omega} \sum_{\gamma \in \Gamma_{x,y}} F[\gamma] k[\gamma] \sum_{i=0}^{k[\gamma]-1} (f(v_{x,y,i+1}) - f(v_{x,y,i}))^2 K(x, y) \mu(x)
\end{aligned}$$

And so the coefficient $[(f(q) - f(r))^2] R_1$ of $(f(q) - f(r))^2$ in R_1 is at most

$$(10) \quad [(f(q) - f(r))^2] R_1 \leq \sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] K(i(\gamma), o(\gamma)) \mu(i(\gamma))$$

The next step is to bound R_2 , which depends on the measures P_x and flow F , though not on the couplings $P_{x,y}$. Write:

$$\begin{aligned}
R_2 &= \sum_{x \in \Omega, y \notin \Omega} |f(x) - \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&= \sum_{x \in \Omega, y \notin \Omega} \left| \sum_{z \in \Omega} P_y[z] (f(x) - f(z)) \right|^2 K(x, y) \mu(x) \\
&\leq \sum_{x \in \Omega, y \notin \Omega} \sum_{z \in \Omega} P_y[z] (f(x) - f(z))^2 K(x, y) \mu(x)
\end{aligned}$$

where the last inequality is Cauchy-Schwarz. The next step is to write $(f(x) - f(z))^2$ in terms of differences which appear in S . To do so, note that

$$\begin{aligned}
(f(x) - f(z))^2 &= \left(\sum_{\gamma \in \Gamma_{x,z}} F[\gamma] \sum_{i=0}^{k[\gamma]-1} (f(v_{x,z,i+1}) - f(v_{x,z,i})) \right)^2 \\
(11) \quad &\leq \sum_{\gamma \in \Gamma_{x,z}} F[\gamma] \left(\sum_{i=0}^{k[\gamma]-1} (f(v_{x,z,i+1}) - f(v_{x,z,i})) \right)^2 \\
&\leq \sum_{\gamma \in \Gamma_{x,z}} F[\gamma] k[\gamma] \sum_{i=0}^{k[\gamma]-1} (f(v_{x,z,i+1}) - f(v_{x,z,i}))^2
\end{aligned}$$

where both inequalities are Cauchy-Schwarz. From this bound, the coefficient of $(f(q) - f(r))^2$ in R_2 is at most

$$(12) \quad [(f(q) - f(r))^2] R_2 \leq \sum_{\gamma \ni (q,r)} k[\gamma] F[\gamma] \sum_{y \notin G} P_y[o(\gamma)] K(i(\gamma), y) \mu(i(\gamma))$$

Finally, it is necessary to bound R_3 . Write

$$\begin{aligned}
R_3 &= \sum_{x,y \in \widehat{\Omega} \setminus \Omega} |\widehat{f}(x) - \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&= \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \left| \sum_{a \in \Omega} P_x[a] f(a) - \sum_{b \in \Omega} P_y[b] f(b) \right|^2 K(x, y) \mu(x) \\
&= \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \left| \sum_{a,b \in \Omega} P_{x,y}[a, b] (f(a) - f(b)) \right|^2 K(x, y) \mu(x) \\
&\leq \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \sum_{a,b \in \Omega} P_{x,y}[a, b] (f(a) - f(b))^2 K(x, y) \mu(x)
\end{aligned}$$

Using inequality (11) above, this gives
(13)

$$R_3 \leq \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \sum_{a,b \in \Omega} P_{x,y}[a, b] \sum_{\gamma \in \Gamma_{a,b}} F[\gamma] k[\gamma] \sum_{i=0}^{k(\gamma)-1} (f(v_{a,b,i+1}) - f(v_{a,b,i}))^2 K(x, y) \mu(x)$$

In particular, the coefficient of $(f(q) - f(r))^2$ in this upper bound is

$$[(f(q) - f(r))^2] R_3 \leq \sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] \sum_{(x,y) \in \widehat{E}, x,y \notin \Omega} P_{x,y}[i(\gamma), o(\gamma)] K(x, y) \mu(x)$$

Combining inequalities (10), (12) and (13), the coefficient of $(f(q) - f(r))^2$ in $R_1 + 2R_2 + R_3$ is bounded by

$$\begin{aligned}
[(f(q) - f(r))^2] (R_1 + 2R_2 + R_3) &\leq \sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] K(i(\gamma), o(\gamma)) \mu(i(\gamma)) \\
&\quad + 2 \sum_{\gamma \ni (q,r)} k[\gamma] F[\gamma] \sum_{y \notin G} P_y[o(\gamma)] K(i(\gamma), y) \mu(i(\gamma)) \\
&\quad + \sum_{\gamma \ni (q,r)} F[\gamma] k[\gamma] \sum_{(x,y) \in \widehat{E}, x,y \notin G} P_{x,y}[i(\gamma), o(\gamma)] K(x, y) \mu(x)
\end{aligned}$$

On the other hand, the coefficient of $(f(q) - f(r))^2$ in $\mathcal{E}_Q(f, f)$ is at least $Q(q, r) \nu(q)$. Thus, setting

$$\begin{aligned}
\mathcal{A} = & \sup_{Q(q,r)>0} \frac{1}{Q(q,r)\nu(q)} \left(\sum_{\gamma \ni (q,r)} F[\gamma]k[\gamma]K(i(\gamma), o(\gamma))\mu(i(\gamma)) \right. \\
& + 2 \sum_{\gamma \ni (q,r)} k[\gamma]F[\gamma] \sum_{y \notin G} P_y[o(\gamma)]K(i(\gamma), y)\mu(i(\gamma)) \\
& \left. + \sum_{\gamma \ni (q,r)} F[\gamma]k[\gamma] \sum_{(x,y) \in \widehat{E}, x,y \notin G} P_{x,y}[i(\gamma), o(\gamma)]K(x, y)\mu(x) \right)
\end{aligned}$$

we have

$$\mathcal{E}_K(\widehat{f}, \widehat{f}) \leq \mathcal{A}\mathcal{E}_Q(f, f)$$

which completes the proof. ■

Theorem 3 is an immediate corollary.

5. SIMPLE EXAMPLES

In this section, we present brief proofs of Theorems 5 and 6, which follow quickly from Theorem 3. We begin with Theorem 5.

Proof. First, we define the measures. If $x \in \{v_1, v_2, \dots, v_m\}$, let

$$P_x[a] = \begin{cases} \frac{1}{4} & \text{if } (x, a) \in \widehat{E} \\ 0 & \text{otherwise} \end{cases}$$

By assumption, no two vertices in $\{v_1, v_2, \dots, v_m\}$ are adjacent, so there are no choices to make when defining the couplings $P_{x,y}$. To define the flow, note that $P_{x,y}[a, b] > 0$ only in three situations. We describe each situation up to swapping coordinates and reflections about rows and columns:

Case 1: $(x, y) = (a, b)$. In this case, $(a, b) \in E$, so define the flow to be concentrated on that single edge.

Case 2: $x = (i, j) \in G$, y of the form $(i + 1, j) \notin G$, $a = x$, and b of the form $(i + 1, j + 1)$. In this case, define the flow to be concentrated on the path $\{((i, j), (i, j + 1)), ((i, j + 1), (i + 1, j + 1))\}$.

Case 3: $x = (i, j) \in G$, y of the form $(i + 1, j) \notin G$, $a = x$, and b of the form $(i + 2, j)$. In this case, there are two length 4 paths between a and b , of the form $\{((i, j), (i, j + 1)), ((i, j + 1), (i + 1, j + 1)), ((i + 1, j + 1), (i + 2, j + 1)), ((i + 2, j + 1), (i + 2, j))\}$ and $\{((i, j), (i, j - 1)), ((i, j - 1), (i + 1, j - 1)), ((i + 1, j - 1), (i + 2, j - 1)), ((i + 2, j - 1), (i + 2, j))\}$. The flow should put equal weight on both.

Then, note that any edge can be in at most 1 path associated with case 1, 2 paths associated with case 2, and 4 paths associated with case 3. Thus, $\mathcal{A} \leq \frac{n-m}{n}(1 + (2)(2)^{\frac{1}{4}} + (4)(4)^{\frac{1}{4}}) = 6(1 - \frac{m}{n})$. ■

As mentioned in the introduction, this bound translates immediately into an $O(n^2 \log(n))$ bound on the Total Variation mixing time, using inequality (7). Unfortunately, although Theorem 5 can be used to get very good control on the entire spectrum of the associated walk as per the comments immediately preceding Lemma 2, the lack of symmetry in the problem makes it difficult to use the smaller eigenvalues to actually improve our estimate of the mixing time. In section 9, we will avoid this problem and find the right bound up to the coefficient of the leading term using the spectral profile.

The proof of Theorem 6 is similar:

Proof. We begin by defining the measures. For $x \in \{v_1, v_2, \dots, v_m\}$, define

$$P_x[a] = \begin{cases} \frac{1}{4} & \text{if } (x, a) \in \hat{E} \\ 0 & \text{otherwise} \end{cases}$$

By assumption, no two vertices in $\{v_1, v_2, \dots, v_m\}$ are adjacent, so it isn't necessary to define any couplings. To define the flow, put the entire weight on one minimal length path. Since the number of pairs (a, b) with a and y not adjacent but $P_{x,y}[a, b] > 0$ for some x, y is at most $2n$, and the maximal path length is clearly at most $4n$, we can write $\mathcal{A} \leq 8n^2$. ■

More importantly, it seems impossible to substantially improve this bound with another comparison to simple random walk on the torus. The missing vertices effectively divide the torus into two regions. There must be at least $\Omega(n)$ paths going between the two regions, and the median path length must also be at least $\Omega(n)$. Since $O(1)$ edges between the two regions exist, any path argument gives $\mathcal{A} = \Omega(n^2)$.

6. THE RANDOM TRANSPOSITION WALK ON DERANGEMENTS

This section contains the proofs of Theorem 7 and Corollary 8. The proof is based on an application of Theorem 3, and the strategy is quite simple. Say that τ is an *extension* of σ if every cycle of σ is contained in some cycle of τ , when cycles are viewed as subsets of $[n]$. Roughly, for $x \in S_n \setminus D_n$, we will define measures P_x supported on D_n which are fairly uniform on derangements that are extensions of x . We will then find a coupling $P_{x,y}[\sigma, \tau]$ of $P_x[\sigma]$ and $P_y[\tau]$ so that if $P_{x,y}[\sigma, \tau] > 0$, there will be a sequence of derangements of length at most 4, starting with σ and ending with τ , where adjacent derangements differ by a single transposition. Finally, the flows will be supported on these minimal-length paths. To complete the proof, we will describe for any fixed pair q, r of derangements all pairs x, y and all pairs σ, τ

so that the edge (q, r) is in a path from σ to τ with $P_{x,y}[\sigma, \tau] > 0$. Most of the details of the proof consist of examining a relatively large number of simple cases. For this reason, we omit some details for cases very similar to those already discussed and instead have only references to the first time the calculation is done.

We begin by setting some notation. Let S_n and D_n be the collection of permutations and derangements on $[n]$, respectively. For $x \in S_n$, define $Fix(x) = \{i \in [n] : x[i] = i\}$ to be the fixed points of x . We will multiply permutations from left to right, so that e.g. $(1, 3)(1, 2) = (1, 3, 2)$. Finally, for $\sigma \in S_n$ and any subset S of n which is exactly the union of cycles of σ , we will denote by $\sigma|_S$ the restriction of σ to S . For example, $(125)(34)|_{\{1,2,5\}} = (125)$. This is often useful for writing down explicit paths along which most such restrictions don't change.

The next step is to describe the measures P_x for $x \in S_n \setminus D_n$, their couplings $P_{x,y}$ for $x, y \in S_n \setminus D_n$, and flows $G_{\sigma,\tau}$ for pairs $\sigma, \tau \in D_n$ with $P_{x,y}[\sigma, \tau] > 0$ for some pair $x, y \in S_n$. We begin with the measures $P_x[\sigma]$. Let $Fix(x) = F = \{f_1, \dots, f_m\}$, and write the remaining cycles of x as $x = C_1 C_2 \dots C_k$, where $C_i = (p_{i,1}, p_{i,2}, \dots, p_{i,\ell(i)})$, with $\ell(i) \geq 2$.

We now write several measures associated with $x \neq id$. Fix $z \in S_m$, and construct σ distributed according to $P_x^{(z)}$ as follows. For $1 \leq i \leq m$, let a_i be chosen uniformly in $[n] \setminus \cup_{j=i}^m \{f_{z(j)}\}$. Then write

$$(14) \quad \sigma = x(a_1, f_{z(1)})(a_2, f_{z(2)}) \dots (a_m, f_{z(m)})$$

Note that the i 'th transposition $(a_i, f_{z(i)})$ is the first time that $f_{z(i)}$ appears in the sequence $(x \prod_{i=1}^j (a_i, f_{z(i)}))_{j=0}^m$, and so the j 'th term in that sequence is obtained from the $j-1$ 'st by adding $f_{z(j)}$ to the cycle containing $a_{z(j)}$. In particular, no cycles are split during this iterative construction. This defines a measure $P_x^{(z)}$ concentrated on D_n . It is worth noting that these measures aren't very uniform. For example, if $Fix(x) = \{a, b\}$ and $P_x[\sigma] > 0$, then (a, b) is not in the cycle decomposition of σ ; if $|Fix(x)| = n-2$, $P_x^{(z)}$ is concentrated on n -cycles. We're willing to give up some uniformity to gain the following lemma, which is very useful for constructing couplings:

Lemma 10 (Order Indifference). *For $x \in S_n \setminus D_n$ with $Fix(x) = \{f_1, f_2, \dots, f_m\} \neq [n]$, $A \subset D_n$, and $z, z' \in S_m$, we have $P_x^{(z)}[A] = P_x^{(z')}[A]$.*

Observe that, under any ordering, $P_x^{(z)}[\sigma] \in \{0, \frac{(n-m-1)!}{(n-1)!}\}$, since each obtainable element can be obtained in a unique way. Next, observe that the supports of $P_x^{(z)}$ and $P_x^{(z')}$ are the same. In cycle notation, they consist of exactly the derangements that can be obtained by slotting the elements of $Fix(x)$ into the non-trivial cycles of x . \square

Using this lemma, define P_x to be the single measure $P_x^{(z)}$ for some (any) ordering z . If $x = id$, then let P_{id} be uniform on n -cycles. These distributions are biased in the sense discussed immediately before the lemma. However, we'll see that they are 'most biased' for permutations with a large number of fixed points, and there aren't enough of those to be significant.

Having described the measures P_x , it is necessary to find couplings $P_{x,y}$ for x, y adjacent. The goal will be to ensure that if $P_{x,y}[\sigma, \tau] > 0$, then the distance between σ and τ is small as measured by the graph metric on D_n induced by the kernel Q . Note that, since x and y are only a transposition away, we can assume without loss of generality that $Fix(x) \subset Fix(y)$.

This observations gives an easy way to construct a coupling, as long as $x, y \neq id$. Under the representation of P_x given by equation (14), we order $Fix(x)$, then order the elements of $Fix(y)$ to put the elements in $Fix(y) \setminus Fix(x)$ at the front, and the remaining elements in the same order as given in $Fix(x)$. Then let $\{a_i\}_{i=1}^{|Fix(y)|}$ be the random variables used to build τ from P_y in representation (14). Construct σ from P_x using the same choices for a_i in representation equation (14) for all $i > |Fix(y)| - |Fix(x)|$. This defines the coupling for $x, y \neq id$. For $y = id$ and $x = (i, j)$, we can observe that $P_{id} = P_{(i,j)}$, so we choose the obvious coupling $P_{id,(i,j)}[\sigma, \tau] = P_{id}(\sigma) \mathbf{1}_{\sigma=\tau}$.

The next step is to define flows between all pairs $\sigma, \tau \in D_n$ such that $P_{x,y}[\sigma, \tau] > 0$ for at least some pair $x, y \in S_n$. We will often use the shorthand "assign weight β to path $\gamma_{\sigma,\tau}$ " for "set $G_{\sigma,\tau}[\gamma_{\sigma,\tau}] = \beta$."

Case 1: $x, y \in D_n$. In this case, $P_{x,y}[\sigma, \tau] = \delta_{x,y}[\sigma, \tau]$. There is an edge between σ and τ , and so we assign weight 1 to that length-1 path.

Case 2: $x \in D_n, |Fix(y)| = 1$. Assume without loss of generality that $Fix(y) = i$. Thus, $x = y(i, j)$ for some j . In this case, $P_{x,y}[\sigma, \tau] > 0$ if and only if $\sigma = x = y(i, j)$ and $\tau = y(i, k)$ for some $k \neq j$. Thus, we need to create a flow from $y(i, j)$ to $y(i, k)$ for all permutations y with unique fixed point i and all distinct $j, k \neq i$. For parity reasons, it is clear that there are no paths of length 1. Let C_j and C_k be the cycles containing j and k respectively in y . By assumption, these have size at least 2, so write $C_j = (j, a_1, \dots, a_{\ell(j)})$ and $C_k = (k, b_1, \dots, b_{\ell(k)})$. There are three subcases to consider:

Case 2A: Assume first that $C_j \neq C_k$. Then $\{(y(i, j), y(i, j)(j, k)), (y(i, j)(j, k), y(i, j)(j, k)(i, j) = y(i, k))\}$ is a length-two path between $y(i, j)$ and $y(i, k)$, and all vertices are clearly in D_n . Assign weight 1 to this path.

Case 2B: $C_j = C_k$, with $|C_j| > 2$. In this case, write $C = C_j = C_k = (j, a_1, \dots, a_{\ell(j)}, k, b_1, \dots, b_{\ell(k)})$. If $\ell(j) > 0$, the path described in case 2A remains in D_n , and so we assign weight 1 to this path. If $\ell(k) > 0$, an analogous path with

j 's and k 's switched and arrows reversed will remain in D_n . Since $|C_j| > 2$, we have either $\ell(j) > 0$ or $\ell(k) > 0$. Thus, the only remaining case is:

Case 2C: $C_j = C_k = (j, k)$. In this case, we will use the assumption that $n \geq 5$. Choose $h \in [n]/\{i, j, k\}$, and write $C_h = (h, c_1, \dots, c_{\ell(h)})$ for some $\ell(h) \geq 1$. Also write $S = \{i, j, k, h, c_1, \dots, c_{\ell(h)}\}$. We calculate:

$$\begin{aligned} y(i, j)|_S &= (i, j, k)(h, c_1, \dots, c_{\ell(h)}) \\ y(i, j)(i, h)|_S &= (i, j, k, h, c_1, \dots, c_{\ell(h)}) \\ y(i, j)(i, h)(j, h)|_S &= (j, k)(h, c_1, \dots, c_{\ell(h)}) \\ y(i, j)(i, h)(j, h)(k, h)|_S &= (i, k, j, h, c_1, \dots, c_{\ell(h)}) \\ y(i, j)(i, h)(j, h)(k, h)(i, h)|_S &= \tau \end{aligned}$$

The five permutations described above, when restricted to S^c , are all equal. These five permutations, without restriction to S , are all in D_n , and so for each $h \in [n]/\{i, j, k\}$ this sequence defines a length-4 path from σ to τ . In this case, we put weight $\frac{1}{n-3}$ on each of these paths from σ to τ .

Case 3: $x \in D_n$, $|Fix(y)| = 2$. Without loss of generality, write $Fix(y) = \{a, b\}$, so that $x = y(a, b)$. Note also that since $x \in D_n$, $P_x[\sigma] = \delta_x[\sigma]$. Write y in cycle notation as $y = (p_{1,1}, \dots, p_{1,\ell(1)})(p_{2,1}, \dots, p_{2,\ell(2)}) \dots (p_{k,1}, \dots, p_{k,\ell(k)})(a)(b)$, where $\ell(i)$ is the length of the i 'th longest cycle, with ties broken lexicographically by smallest element. If $P_y[\tau] > 0$, we can write $\tau = y(a, p_{i(a),j(a)})(b, p_{i(b),j(b)})$ or $\tau = y(a, p_{i(a),j(a)})(b, a)$, where in both cases $1 \leq i(a), i(b) \leq k$, $1 \leq j(a) \leq \ell(i(a))$, $1 \leq j(b) \leq \ell(i(b))$. This leads to three types of paths.

Case 3A: $\tau = y(a, p_{i(a),j(a)})(b, p_{i(b),j(b)})$ with $i(a) \neq i(b)$. We define two paths from τ to σ as follows. In both paths, the first derangement is τ . The second is given by either $\tau(b, p_{i(a),j(a)})$ or $\tau(a, p_{i(b),j(b)})$. The symmetry between these two first steps being clear, we continue describing only the path beginning $(\tau, \tau(b, p_{i(a),j(a)}), \dots)$. Note that the cycle structure of $\tau(b, p_{i(a),j(a)})$ is given by the cycle structure of τ with the two cycles $(p_{i(a),1}, \dots, p_{i(a),j(a)-1}, a, p_{i(a),j(a)}, \dots, p_{i(a),\ell(i(a))})$ and $(p_{i(b),1}, \dots, p_{i(b),j(b)-1}, b, p_{i(b),j(b)}, \dots, p_{i(b),\ell(i(b))})$ merged into the single cycle

$$(p_{i(a),j(a)}, p_{i(a),j(a)+1}, \dots, p_{i(a),j(a)-1}, a, b, p_{i(b),j(b)}, \dots, p_{i(b),j(b)-1})$$

In particular, it is still a derangement. The next step on this path is $\tau(b, p_{i(a),j(a)})(a, p_{i(b),j(b)})$. The cycle structure of this permutation is obtained from that of $\tau(b, p_{i(a),j(a)})$ by splitting the large cycle $(p_{i(a),j(a)}, p_{i(a),j(a)+1}, \dots, p_{i(a),j(a)-1}, a, b, p_{i(b),j(b)}, \dots, p_{i(b),j(b)-1})$ into the smaller cycles (a, b) and $(p_{i(a),j(a)}, p_{i(a),j(a)+1}, \dots, p_{i(a),j(a)-1}, p_{i(b),j(b)}, \dots, p_{i(b),j(b)-1})$. Again, this is a derangement. The final step is multiplying by $(p_{i(b),j(b)}, p_{i(a),j(a)})$ to get to σ . We assign weight $\frac{1}{2}$ to both paths.

Case 3B: $\tau = y(a, p_{i(a),j(a)})(b, p_{i(b),j(b)})$ with $i(a) = i(b)$ and $\tau(a) \neq b, \tau(b) \neq a$. We create the following two paths from τ to σ . The first vertex is τ . The second vertex is $\tau(a, b)$, which has the same cycle structure as τ , with the long cycle $(p_{i(a),j(a)-1}, a, p_{i(a),j(a)}, \dots, p_{i(a),j(b)-1}, b, p_{i(a),j(b)}, \dots)$ split into the cycles $(a, p_{i(a),j(a)}, \dots, p_{i(a),j(b)-1})$ and $(b, p_{i(a),j(b)}, \dots, p_{i(a),j(a)-1})$. By the assumption that a and b were not adjacent in the large cycle, both of the small cycles are of size at least 2, so this is a derangement. The next vertex should be either $\tau(a, b)(p_{i(a),j(a)}, b)$ or $\tau(a, b)(p_{i(a),j(b)}, a)$. As in case 3A, there is obvious symmetry after relabelling a and b , and we will continue the description of the first of these paths. Note that $\tau(a, b)(p_{i(a),j(a)}, b)$ obtained from $\tau(a, b)$ by merging the cycles $(a, p_{i(a),j(a)}, \dots, p_{i(a),j(b)-1})$ and $(b, p_{i(a),j(b)}, \dots, p_{i(a),j(a)-1})$ into the single cycle $(a, b, p_{i(a),j(b)}, \dots, p_{i(a),j(a)-1}, p_{i(a),j(a)}, \dots, p_{i(a),j(b)-1})$. This is clearly a derangement. Finally, send $\tau(a, b)(p_{i(a),j(a)}, b)$ to σ by multiplying by the transposition $(a, p_{i(a),j(b)})$. The path with the other middle edge is analogous; we assign weight $\frac{1}{2}$ to both paths.

Case 3C: This covers the cases $\sigma = y(a, p_{i(a),j(a)})(b, a)$ and $\tau = y(a, p_{i(a),j(a)})(b, p_{i(b),j(b)})$ with $i(a) = i(b)$ and either $\tau(a) = b$ or $\tau(b) = a$. In this case, σ is adjacent to τ , and in particular $\sigma = \tau(p_{i(a),j(a)}, b)$. The flow should put all weight on this length-1 path.

The next step is to look at the cases where $x, y \in S_n \setminus D_n$. There will be 3 cases, depending on whether $|Fix(x)| = |Fix(y)|$, $|Fix(x)| = |Fix(y)| - 1$, or $|Fix(x)| = |Fix(y)| - 2$. These will turn out to be very similar to cases 1 through 3 above, with slightly more complicated notation. In particular, all paths will again be of length at most 4.

Case 4: $|Fix(x)| = |Fix(y)|$. In this case, we can assume without loss of generality that y has the same cycle structure as x with the i 'th cycle, $(p_{i,1}, \dots, p_{i,\ell(i)})$, split into the two cycles $(p_{i,1}, \dots, p_{i,a})$ and $(p_{i,a+1}, \dots, p_{i,\ell(i)})$. Let $Fix(x) = Fix(y) \equiv F = \{f_1, \dots, f_m\}$. If $P_{x,y}[\sigma, \tau] > 0$, then σ and τ have the same cycle structure, except that the single cycle including $p_{i,1}, \dots, p_{i,\ell(i)}$ in that order in σ is split into two cycles in τ . One contains $p_{i,1}, \dots, p_{i,a}$ in that order and the other contains $p_{i,a+1}, \dots, p_{i,\ell(i)}$ in that order. Both will have some elements of F interspersed between the elements of the form $p_{i,q}$, but these interspersed elements will also always be in the same order in σ and τ .

In particular, for some $0 \leq \alpha \leq m$, some $z \in S_m$ and $\phi : [\alpha] \rightarrow \ell(i)$ we can write $S = \{p_{i,1}, \dots, p_{i,\ell(i)}\} \cup \{f_{z(1)}, \dots, f_{z(\alpha)}\}$ and

$$\begin{aligned}\sigma|_S &= x \prod_{k=1}^{\alpha} (p_{i,\phi(k)}, f_{z(k)}) \\ \tau|_S &= x(p_{i,1}, p_{i,a+1}) \prod_{k=1}^{\alpha} (p_{i,\phi(k)}, f_{z(k)})\end{aligned}$$

It is easy to check that σ and τ are adjacent, and in fact $\sigma = \tau(\tau[p_{i,a}], \tau[p_{i,\ell(i)}])$. Assign weight 1 to this length-1 path.

Case 5: $|Fix(x)| = |Fix(y)| - 1$. Assume without loss of generality that $Fix(x) = \{f_1, \dots, f_m\}$ and $Fix(y) = \{i, f_1, \dots, f_m\}$. Therefore, $x = y(i, j)$ for some $j \neq i$, and $P_{x,y}[\sigma, \tau] > 0$ if and only if σ, τ can be written in the form

$$(15) \quad \sigma = y(i, j) \prod_{s=1}^m (a_s, f_s)$$

$$(16) \quad \tau = y(i, k) \prod_{s=1}^m (a_s, f_s)$$

with $a_s \in [n] \setminus \cup_{t \geq s} \{f_t\}$. We will construct paths very similar to those in Case 2. For $\alpha \in [n]$, let $F(\alpha) = (f_1(\alpha), \dots, f_{\ell(\alpha)}(\alpha))$ subset of F with the property that $f_s \in F$ if $a_s = \alpha$ or $a_s \in F$, using the representation (15). We order $F(\alpha)$ by the indices, so that if $f_i, f_j \in F(\alpha)$ with $i < j$, then f_i is before f_j in the list $F(\alpha)$. When $F(\alpha)$ is empty, define $f_1(\alpha) = \alpha$.

Next, say that a directed edge (η, κ) in S_n is *defined by the transposition* (q, r) if $\eta^{-1}\kappa = (q, r)$. We will be changing paths by changing the transpositions that define their edges. In general, if $\gamma_{\sigma, \tau} = (\sigma, \sigma(q_1, r_1), \dots, \sigma \prod_{i=1}^k (q_i, r_i) = \tau)$ is a path from σ to τ , we say that $\gamma_{\sigma', \tau'} = (\sigma', \sigma'(q'_1, r'_1), \dots, \sigma' \prod_{i=1}^k (q'_i, r'_i) = \tau')$ is the path from σ' to τ' obtained by replacing all edges defined by (q_i, r_i) to edges defined by (q'_i, r'_i) . To define the flows in case 5, we will take paths from case 2 and replace all edges defined by transpositions (q, r) with edges defined by transposition $(f_1(q), f_1(r))$.

We will say this more carefully for the analogue to case 2A. Assume j, k aren't in the same cycle in y , and let S be the union of all elements in cycles containing i, j or k in σ or τ . Then

$$\begin{aligned}\sigma|_S &= (F(i), i, F(j), j, b_1, \dots, b_{\ell(j)})(F(k), k, c_1, \dots, c_{\ell(k)}) \\ \tau|_S &= (F(i), i, F(k), k, c_1, \dots, c_{\ell(k)})(F(j), j, b_1, \dots, b_{\ell(j)})\end{aligned}$$

And so we write the path $\{(\sigma, \sigma(f_1(j), f_1(k))), (\sigma(f_1(j), f_1(k)), \sigma(f_1(j), f_1(k))(f_1(j), f_1(i)))\}$. This replaces the analogous path $\{(\sigma, \sigma(j, k)), (\sigma(j, k), \sigma(j, k)(i, j))\}$ from case 2A. As in case 2B, we will use the same path if j, k are in the same cycle in y and $(i, j, k) \notin \sigma, \tau$.

If $(i, j, k) \in \sigma, \tau$, the same discussion as in case 2C shows that the path which goes through σ , $\sigma(f_1(i), f_1(h)), \sigma(f_1(i), f_1(h))(f_1(j), f_1(h)), \sigma(f_1(i), f_1(h))(f_1(j), f_1(h))(f_1(k), f_1(h))$ and finally $\sigma(f_1(i), f_1(h))(f_1(j), f_1(h))(f_1(k), f_1(h))(f_1(i), f_1(h)) = \tau$ remains in D_n .

Case 6: $|Fix(x)| = |Fix(y)| - 2$. Just as case 5 is very similar to case 2, case 6 is very similar to case 3. Assume $Fix(y) = \{f_1, \dots, f_m\}$ and $Fix(x) = \{c, d, f_1, \dots, f_m\}$. In particular, $y = x(c, d)$. Write y in cycle notation, as in case 0c. Then if $P_{x,y}[\sigma, \tau] > 0$, analogously to case 3, we can write the pair (σ, τ) in one of the following three ways:

Case 6A:

$$\begin{aligned}\tau &= y \prod_{s=1}^m (a_s, f_s) \\ \sigma &= y(c, p_{i(c), j(c)})(d, p_{i(d), j(d)}) \prod_{s=1}^m (a_s, f_s)\end{aligned}$$

with $i(c) \neq i(d)$,

Case 6B:

$$\begin{aligned}\tau &= y \prod_{s=1}^m (a_s, f_s) \\ \sigma &= y(c, p_{i(c), j(c)})(d, p_{i(d), j(d)}) \prod_{s=1}^m (a_s, f_s)\end{aligned}$$

with $i(c) = i(d)$, or

Case 6C:

$$\begin{aligned}\tau &= y \prod_{s=1}^m (a_s, f_s) \\ \sigma &= y(c, p_{i(c),j(c)})(d, c) \prod_{s=1}^m (a_s, f_s)\end{aligned}$$

where in each case $a_s \in [n] \setminus \cup_{t \geq s} \{f_t\}$. These three possibilities correspond exactly to those in cases 3A, 3B and 3C respectively. Just as in case 5, we define flows by taking the paths in cases 3A, 3B and 3C and substituting an edge defined by transposition $(f_1(q), f_1(r))$ for any edge defined by transposition (q, r) in case 3.

Having defined the measures, couplings, and flows, we will now bound the comparison constant \mathcal{A} . We will do this by bounding separately the edges that appear in paths between σ and τ with $P_{x,y}[\sigma, \tau] > 0$ and x, y satisfying the conditions from cases 1 through 6 above. To be more precise, for $\mathcal{F} \in \{1, 2, \dots, 6\}$ and $x, y \in S_n$, say that $(x, y) \in \mathcal{F}$ if $x^{-1}y$ is a transposition and x, y satisfies the conditions of case \mathcal{F} above. We then write:

$$\begin{aligned}\mathcal{A} &= \sup_{(q,r) \in E} (1 + 2 \sum_{\gamma_{x,z} \ni (q,r)} k[\gamma_{x,z}] G_{x,z}[\gamma_{x,z}] \sum_{y \notin G} P_y[z] \\ &\quad + \sum_{\gamma_{a,b} \ni (q,r)} G_{a,b}[\gamma_{a,b}] k[\gamma_{a,b}] \sum_{(x,y) \in \hat{E}, x,y \notin G} P_{x,y}[a, b]) \\ &\leq 1 + \sup_{(q,r)} \sum_{\mathcal{F}} \sum_{\gamma_{x,z} \ni (q,r), (x,z) \in \mathcal{F}} k[\gamma_{x,z}] G_{x,z}[\gamma_{x,z}] \sum_{y \notin G} P_y[z] \\ &\quad + \sup_{(q,r)} \sum_{\mathcal{F}} \sum_{\gamma_{a,b} \ni (q,r)} G_{a,b}[\gamma_{a,b}] k[\gamma_{a,b}] \sum_{(x,y) \in \hat{E}, x,y \notin G, (x,y) \in \mathcal{F}} P_{x,y}[a, b])\end{aligned}$$

We will then separately bound the weights associated with each of the 6 cases in this sum. In principle, this part of the argument is the same as the weight-counting at the end of the proof of Theorem 5; it is only the larger number of terms in each case that makes it more complicated.

Case 1: $(q, r) \in \gamma_{\sigma, \tau}$, where $G_{\sigma, \tau}[\gamma_{\sigma, \tau}] > 0$ and $P_{x,y}[\sigma, \tau] > 0$ for some $x, y \in D_n$. This implies that in fact $(q, r) = (x, y)$, and so the total weight in this case is exactly 1.

Case 2: $(q, r) \in \gamma_{\sigma, \tau}$, where $G_{\sigma, \tau}[\gamma_{\sigma, \tau}] > 0$ and $P_{x, y}[\sigma, \tau] > 0$ for some $x \in D_n$, $|Fix(y)| = 1$. Assume $Fix(y) = i$. Then $x = y(i, j)$ for some j , and we can write $\sigma = x = y(i, j)$, $\tau = y(i, k)$. If x, y correspond to case 2A or 2B above, the path between σ and τ passes through the permutations $\{y(i, j), y(i, j)(j, k), y(i, j)(j, k)(i, j)\}$. Thus, either $(q, r) = (y(i, j), y(i, j)(j, k))$ or $(q, r) = (y(i, j)(j, k), y(i, j)(j, k)(i, j))$.

First, assume $(q, r) = (y(i, j), y(i, j)(j, k))$. Then $q(i, j)$ and $r(j, k)(i, j)$ have a fixed point at i . This means that $q[j] = i$ and $r[k] = i$. In particular, once q, r and i have been fixed, so are j and k . Thus, for fixed q, r there are at most $n - 1$ choices for i , and these choices determine x, y, σ and τ . Since $P_{x, y}[\sigma, \tau] \in \{0, \frac{1}{n-1}\}$, this means that the total weight is at most $(n - 1)\frac{1}{n-1} = 1$.

If $(q, r) = (y(i, j)(j, k), y(i, j)(j, k)(i, j))$, a similar computation gives the same conclusion. Thus, the total weight for any given edge (q, r) coming from pairs x, y in case 2A or 2B is at most 2.

If x, y correspond to case 2C, there are four possibilities for the pair (q, r) , as described in case 2C. Following the notation in that case, we look at the first possibility, $(q, r) = (y(i, j), y(i, j)(i, h))$. Note that $q[i] = j$ and $r[(q(i, j))^{-1}[j]] = h$. In particular, once q, r, k and i have been fixed, $j = q[i]$ can be computed from them, and this information can be used to compute $h = r[(q(i, j))^{-1}[j]]$. Thus, for fixed q, r there are at most $(n - 1)(n - 2)$ choices of distinct i, k , and these choices determine x, y, σ and τ . Since $P_{x, y}[\sigma, \tau] \in \{0, \frac{1}{n-1}\}$ and the weight of any particular path in case 2C is $\frac{1}{n-3}$, the total weight assigned to any first edge (q, r) by such a path is at most $\frac{n-2}{n-3}$. A similar analysis with the same conclusion applies to the other 3 edges of the length-4 paths described in case 2C.

We conclude that the total weight assigned to any edge (q, r) by vertices (x, y) covered by case 2 is at most $6\frac{n-2}{n-3}$.

Case 3: $(q, r) \in \gamma_{\sigma, \tau}$, where $G_{\sigma, \tau}[\gamma_{\sigma, \tau}] > 0$ and $P_{x, y}[\sigma, \tau] > 0$ for some $x \in D_n$, $|Fix(y)| = 2$. Write in this case $Fix(y) = \{a, b\}$, so $x = y(a, b)$. There are three possibilities for pairs (σ, τ) with $P_{x, y}[\sigma, \tau] > 0$, corresponding to cases 3A, 3B and 3C. We keep the same cycle notation as in case 3 above and begin by looking at case 3A. In this case, the path from τ to σ has the form

$$\begin{aligned} \tau &\rightarrow \tau(b, p_{i(a), j(a)}) \\ &\rightarrow \tau(b, p_{i(a), j(a)})(a, p_{i(b), j(b)}) \\ &\rightarrow \tau(b, p_{i(a), j(a)})(a, p_{i(b), j(b)})(p_{i(b), j(b)}, p_{i(a), j(a)}) = \sigma \end{aligned}$$

Note that the pair (q, r) can be any of the 3 edges defined by this path. Look for now at edges of the form $(q, r) = (\tau, \tau(b, p_{i(a), j(a)}))$. In this case, since $\tau = y(a, p_{i(a), j(a)})(b, p_{i(b), j(b)})$, where y has fixed points at a and b , we note that $q(b, p_{i(b), j(b)})(a, p_{i(a), j(a)})$ and $r(b, p_{i(a), j(a)})(b, p_{i(b), j(b)})(a, p_{i(a), j(a)})$ also have fixed points at a and b . In particular, $q[a] = p_{i(a), j(a)}$, $q[b] = p_{i(b), j(b)}$, $r[a] = b$ and $r[b] =$

$p_{i(b),j(b)}$. Thus, once q, r, a , and b have been fixed, they determine $(p_{i(a),j(a)}, p_{i(b),j(b)}, \sigma$, and τ . For fixed q, r , there are at most $\frac{n(n+1)}{2}$ choices of (a, b) . On the other hand, for a given x, y , the probability $P_{x,y}[\sigma, \tau]$ for any pair σ, τ associated with a, b is at most $\frac{1}{(n-1)(n-2)}$. Thus, the total weight assigned to this path is at most $\frac{n(n+1)}{2(n-1)(n-2)}$. Comparing the pair (q, r) to the remaining two edges of the path, the same phenomenon holds: the fact that y has two fixed points means that the choice of two parameters determines the entire path, and so again the weight given to these edges is at most $\frac{n(n+1)}{2(n-1)(n-2)}$. Combining these bounds, we see that this case gives a total weight of at most $3 \frac{n(n+1)}{2(n-1)(n-2)}$.

Looking at the second case, $\tau = y(a, p_{i(a),j(a)})(b, p_{i(b),j(b)})$ with $i(a) = i(b)$, gives the same congestion bound of $3 \frac{n(n+1)}{2(n-1)(n-2)}$ with essentially the same proof. The third case, $\tau = y(a, p_{i(a),j(a)})(a, b)$, is essentially the same as case 1. As in that case, we have σ and τ adjacent and again determined by the choice of (a, b) . Thus, the total congestion in this case is at most 1 for $n > 6$.

Putting these bounds together, the total weight for any given edge (q, r) coming from pairs (x, y) in this case is at most $1 + 6 \frac{n(n+1)}{2(n-1)(n-2)}$.

Case 4: $(q, r) \in \gamma_{\sigma, \tau}$, where $G_{\sigma, \tau}[\gamma_{\sigma, \tau}] > 0$ and $P_{x,y}[\sigma, \tau] > 0$ for some $x, y \notin D_n$, $|Fix(x)| = |Fix(y)|$. As noted in the coupling description for case 4, this means $(q, r) = (\sigma, \tau)$. We now determine the total weight given to the pair (σ, τ) by all pairs x, y with $|Fix(x)| = j$.

First, note that for any particular pair x, y with $|Fix(x)| = j$, we have $P_{x,y}[\sigma, \tau] \leq \frac{(j-1)!}{(n-1)!}$. Next, note that there are at most $\frac{n!}{j!(n-j)!}$ such pairs x, y for which $P_{x,y}[\sigma, \tau] > 0$; we obtain the bound by noting this is the number of ways to choose the j elements of $Fix(x)$. Summing over the size j of $Fix(x) = Fix(y)$, the total weight assigned to σ, τ is at most

$$\begin{aligned}
W &\equiv \sum_{j=1}^{n-2} \frac{n!}{j!(n-j)!} \frac{(j-1)!}{(n-1)!} \\
&= \sum_{j=1}^{n-2} \frac{n}{j} \frac{1}{(n-j)!} \\
&\leq \sum_{j=1}^{\frac{n}{2}} n \frac{1}{(\frac{n}{2})!} + \sum_{j=\frac{n}{2}}^{n-2} \frac{2}{(n-j)!} \\
&\leq \frac{n^2}{(\frac{n}{2})!} + 2 \sum_{j=2}^{\infty} \frac{1}{j!} \\
&\leq 1 + 2(e-1)
\end{aligned}$$

where the last inequality only applies for n sufficiently large that $n^2 \leq \frac{n!}{2}$, e.g. $n > 10$ suffices. Thus, the total weight for any given edge (q, r) coming from pairs (x, y) in this case is at most $2e + 1$.

Case 5: $(q, r) \in \gamma_{\sigma, \tau}$, where $G_{\sigma, \tau}[\gamma_{\sigma, \tau}] > 0$ and $P_{x, y}[\sigma, \tau] > 0$ for some $x, y \notin D_n$, $|Fix(y)| = |Fix(x)| - 1$. The counting of paths for fixed $|Fix(y)|$ is as in case 2, and finding the weights of each path and summing is as case 4. More precisely, as in case 4, if $|Fix(y)| = j$, the weight assigned to (σ, τ) by $P_{x, y}$ is at most $\frac{(j-1)!}{(n-1)!}$. By the same argument as in case 2, the total weight for any edge $(q, r) \in \gamma_{\sigma, \tau}$ is at most $6 \frac{n-2}{n-3}$, and there are at most $\frac{n!}{j!(n-j)!}$ pairs (x, y) for which $P_{x, y}[\sigma, \tau] > 0$. Thus, a calculation analogous to the bound on W in case 4 gives a total weight of at most $6(2e+1) \frac{n-2}{n-3}$. Combining these arguments gives a total weight of at most $6(2e+1) \frac{n-2}{n-3}$.

Case 6: $(q, r) \in \gamma_{\sigma, \tau}$, where $G_{\sigma, \tau}[\gamma_{\sigma, \tau}] > 0$ and $P_{x, y}[\sigma, \tau] > 0$ for some $x, y \notin D_n$, $|Fix(y)| = |Fix(x)| + 2$. Combining the arguments of 4 and 3 in the same way that case 5 combined the arguments of cases 4 and 2 gives a total weight of at most $(2e+1)(1 + 3 \frac{n(n+1)}{(n-1)(n-2)})$.

Putting together the 6 bounds, and noting that all paths are of length at most 4, we have $\mathcal{A} \leq 2(e+1) \left(2 + 6 \frac{n-2}{n-3} + 3 \frac{n(n+1)}{(n-1)(n-2)} \right)$. This proves the upper bound in Theorem 7.

To prove the lower bound, define for $0 \leq \epsilon \leq 1$ the distribution π_ϵ on S_n by $\pi_\epsilon(\sigma) = Z$ for $\sigma \in D_n$ and $\pi_\epsilon(\sigma) = Z\epsilon$ for $\sigma \in S_n \setminus D_n$, where $\frac{1}{n!} \leq Z \leq \frac{2e}{n!}$ is a normalizing constant. Then define the kernel Q_ϵ to be the Metropolis kernel associated with base chain Q and stationary distribution π_ϵ . Let $1 = \beta_{\epsilon, 0} \geq \beta_{\epsilon, 1} \geq \dots \geq \beta_{\epsilon, n!} \geq 0$ be the

spectrum of Q_ϵ . Note that $Q_1 = Q$, and $Q_0 = K$ when restricted to D_n . By Cauchy's interlacing theorem, if we denote by β_1 the second-largest eigenvalue of K , we have $1 - \beta_{0,1} \leq 1 - \beta_1$.

Next, we compare Q_ϵ and $Q_{\epsilon'}$ for $\epsilon > \epsilon' > 0$. Since $Q_\epsilon[\sigma, \tau] > 0$ if and only if $Q_{\epsilon'}[\sigma, \tau] > 0$, all paths can be made length 1. For this choice of paths, there are three types of edges to look at: those between two elements of D_n , those between an element of D_n and an element of $S_n \setminus D_n$, and finally those between two elements of $S_n \setminus D_n$. In all three cases, the congestion and path length is 1. In the first case, $\frac{1}{\pi_\epsilon(z)Q_\epsilon(z,w)}\pi_{\epsilon'}(z)Q_\epsilon(z,w) \leq 2e$. In the second case, $\frac{1}{\pi_\epsilon(z)Q_\epsilon(z,w)}\pi_{\epsilon'}(z)Q_\epsilon(z,w) \leq 2e\frac{\epsilon'}{\epsilon} \leq 2e$. In the third case, $\frac{1}{\pi_\epsilon(z)Q_\epsilon(z,w)}\pi_{\epsilon'}(z)Q_\epsilon(z,w) \leq 2e\left(\frac{\epsilon'}{\epsilon}\right)^2 \leq 2e$. Thus, $\epsilon_{Q_\epsilon}(f, f) \leq 2e\epsilon_{Q_1}(f, f)$ for all $\epsilon > 0$. Taking the limit as ϵ goes to 0, $\epsilon_{Q_0}(f, f) \leq 2e\epsilon_{Q_1}(f, f)$. Again, by Cauchy's interlacing theorem, $\epsilon_K(f, f) \leq 2e\epsilon_{Q_1}(\hat{f}, \hat{f})$. This completes the proof of Theorem 7. \square

To prove Corollary 8, we note that

$$\begin{aligned} 1 - \beta_1(Q) &= \min_{f \neq 0} \frac{\mathcal{E}_Q(f, f)}{V_\nu(f)} \\ &\geq \min_{f \neq 0} 132(1 + e) \frac{\mathcal{E}_K(\hat{f}, \hat{f})}{V_\mu(\hat{f})} \\ &\geq 264(1 + e) \frac{1}{n} \end{aligned}$$

Where the first inequality is due to Lemma 2, Theorem 7 and the fact that $\frac{|S_n|}{|D_n|} \leq 3$ for $n \geq 10$, and the second inequality is due to the spectral gap estimate in section 9.2 of [22]. The bound on $\alpha(Q)$ is found the same way, and relies on the bound $\alpha(K) = \Omega(\frac{1}{n \log(n)})$ found in [17].

7. REVERSIBILITY AND LAZINESS ASSUMPTIONS

Theorem 4 doesn't make use of either the reversibility or $\frac{1}{2}$ -laziness assumptions. However, in order to obtain Total Variation mixing time bounds, we used Theorem 1, which uses both assumptions. In this section, we briefly discuss the common techniques for avoiding these assumptions. To use comparison without reversibility, Cheeger's inequality is often used. In particular, section 4 of [10] applies without modification to the setting of this paper.

Avoiding the laziness assumption requires slightly more work, but is more effective. For non-lazy chains, the term $\beta_1(P)$ in Theorem 1 may be replaced by $\max(\beta_1(P), |\beta_{|X|}(P)|)$. To estimate $|\beta_{|X|}(P)|$, define the following analogue to the Dirichlet form:

$$\begin{aligned}
\mathcal{F}_P(f, f) &= \langle (I + P)f, f \rangle \\
&= \frac{1}{2} \sum_{x, y \in X} |f(x) + f(y)|^2 P(x, y) \pi(x)
\end{aligned}$$

By inequality 2.3 of [5], if $\mathcal{F}_K(\widehat{f}, \widehat{f}) \leq \mathcal{A}\mathcal{F}_Q(f, f)$ and $C = \sup_{y \in \Omega} \frac{\nu(y)}{\mu(y)}$, we have

$$\begin{aligned}
\beta_1(Q) &\geq \beta_{|\Omega|}(Q) \\
&\geq -1 + \frac{C}{\mathcal{A}}(1 + \beta_{|\Omega|}(K)) \\
&\geq -1 + \frac{C}{\mathcal{A}}(1 + \beta_{|\widehat{\Omega}|}(K))
\end{aligned}$$

As per the comments immediately following Lemma 2, it is possible to obtain bounds on more eigenvalues if there is a more structured relationship between f, \widehat{f} .

Define paths, flows, extensions, and couplings as in the proof of Theorem 4, with the added requirement that flows must be concentrated on paths γ with $|\gamma|$ odd. For a given edge e in path γ , let $t_e(\gamma)$ be the number of times that e is traversed in γ . In Theorem 4, we could assume without loss of generality that this was at most 1; in the present context, we can assume that it is at most 2. Then we have the following comparison result:

Theorem 11 (Comparison of Forms for General Chains). *For flows, distributions and couplings as described above,*

$$\mathcal{F}_K(\widehat{f}, \widehat{f}) \leq \mathcal{A}\mathcal{F}_Q(f, f)$$

where

$$\begin{aligned}
\mathcal{A} &= \sup_{Q(q, r) > 0} \frac{1}{Q(q, r)\nu(q)} \left(\sum_{\gamma_{x, y} \ni (q, r)} G_{x, y}(\gamma) t_{(q, r)}(\gamma) k[\gamma] K(x, y) \mu(x) \right. \\
&\quad + 2 \sum_{\gamma_{x, z} \ni (q, r)} t_{(q, r)}(\gamma) k[\gamma_{x, z}] G_{x, z}[\gamma_{x, z}] \sum_{y \notin G} P_y[z] K(x, y) \mu(x) \\
&\quad \left. + \sum_{\gamma_{a, b} \ni (q, r)} G_{a, b}[\gamma_{a, b}] t_{(q, r)}(\gamma) k[\gamma_{a, b}] \sum_{(x, y) \in \widehat{E}, x, y \notin G} P_{x, y}[a, b] K(x, y) \mu(x) \right)
\end{aligned}$$

Proof. Start by writing

$$\begin{aligned}
\mathcal{F}_K(\widehat{f}, \widehat{f}) &= \frac{1}{2} \sum_{x,y \in \widehat{\Omega}} |\widehat{f}(x) + \widehat{f}(y)|^2 K(x,y) \mu(x) \\
&= \frac{1}{2} \sum_{x,y \in \Omega} |f(x) + f(y)|^2 K(x,y) \mu(x) + \sum_{x \in \Omega, y \notin \Omega} |f(x) + \widehat{f}(y)|^2 K(x,y) \mu(x) \\
&\quad + \frac{1}{2} \sum_{x,y \notin \Omega} |\widehat{f}(x) + \widehat{f}(y)|^2 K(x,y) \mu(x) \\
&\equiv \frac{1}{2} R_1 + R_2 + \frac{1}{2} R_3
\end{aligned}$$

The goal is to compare this to $\mathcal{F}_Q(f, f) = \frac{1}{2} \sum_{x,y \in \Omega} |f(x) + f(y)|^2 Q(x,y) \nu(x)$. We begin by looking at R_1 (note that the assumption of odd path length occurs on the second line):

$$\begin{aligned}
R_1 &= \sum_{x,y \in \Omega} |f(x) + f(y)|^2 K(x,y) \mu(x) \\
&= \sum_{x,y \in \Omega} \left| \sum_{\gamma \in \Gamma_{x,y}} G_{x,y}(\gamma) \sum_{i=0}^{k[\gamma]-1} (-1)^i (f(v_{x,y,i+1}) + f(v_{x,y,i})) \right|^2 K(x,y) \mu(x) \\
&\leq \sum_{x,y \in \Omega} \sum_{\gamma \in \Gamma_{x,y}} G_{x,y}(\gamma) \left| \sum_{i=0}^{k[\gamma]-1} (-1)^i (f(v_{x,y,i+1}) + f(v_{x,y,i})) \right|^2 K(x,y) \mu(x) \\
&\leq \sum_{x,y \in \Omega} \sum_{\gamma \in \Gamma_{x,y}} G_{x,y}(\gamma) k[\gamma] \sum_{i=0}^{k[\gamma]-1} (f(v_{x,y,i+1}) + f(v_{x,y,i}))^2 K(x,y) \mu(x)
\end{aligned}$$

And so the coefficient of $(f(q) + f(r))^2$ in R_1 is at most

$$(17) \quad [(f(q) + f(r))^2] R_1 \leq \sum_{\gamma_{x,y} \ni (q,r)} G_{x,y}(\gamma) t_{(q,r)}(\gamma) k[\gamma] K(x,y) \mu(x)$$

The next step is to bound R_2 , which depends on the measures P_x and flows $G_{x,y}$, though not on the couplings $P_{x,y}$. Write:

$$\begin{aligned}
R_2 &= \sum_{x \in \Omega, y \notin \Omega} |f(x) + \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&= \sum_{x \in \Omega, y \notin \Omega} \left| \sum_{z \in \Omega} P_y[z] (f(x) + f(z)) \right|^2 K(x, y) \mu(x) \\
&\leq \sum_{x \in \Omega, y \notin \Omega} \sum_{z \in \Omega} P_y[z] (f(x) + f(z))^2 K(x, y) \mu(x) \\
&\leq \sum_{x \in \Omega, y \notin \Omega} \sum_{z \in \Omega} P_y[z] \sum_{\gamma \in \Gamma_{x,z}} G_{x,z}(\gamma) k[\gamma] \sum_{i=0}^{k[\gamma]-1} (f(v_{x,z,i+1}) + f(v_{x,z,i}))^2 K(x, y) \mu(x)
\end{aligned}$$

where the last inequality is Cauchy-Schwarz. The next step is to write $(f(x) + f(z))^2$ in terms of differences which appear in S . To do so, note that

$$\begin{aligned}
(f(x) + f(z))^2 &= \left(\sum_{\gamma \in \Gamma_{x,z}} G_{x,z}(\gamma) \sum_{i=0}^{k(\gamma)-1} (-1)^i (f(v_{x,z,i+1}) + f(v_{x,z,i})) \right)^2 \\
(18) \quad &\leq \sum_{\gamma \in \Gamma_{x,z}} G_{x,z}(\gamma) \left(\sum_{i=0}^{k(\gamma)-1} (-1)^i (f(v_{x,z,i+1}) + f(v_{x,z,i})) \right)^2 \\
&\leq \sum_{\gamma \in \Gamma_{x,z}} G_{x,z}(\gamma) k[\gamma] \sum_{i=0}^{k(\gamma)-1} (f(v_{x,z,i+1}) + f(v_{x,z,i}))^2
\end{aligned}$$

where both inequalities are just Cauchy-Schwarz. From this bound, the coefficient of $(f(q) + f(r))^2$ in R_2 is at most

$$(19) \quad [(f(q) + f(r))^2] R_2 \leq \sum_{\gamma_{x,z} \ni (q,r)} t_{(q,r)}(\gamma) k[\gamma_{x,z}] G_{x,z}[\gamma_{x,z}] \sum_{y \notin G} P_y[z] K(x, y) \mu(x)$$

Finally, it is necessary to bound R_3 . Write

$$\begin{aligned}
R_3 &= \sum_{x,y \in \widehat{\Omega} \setminus \Omega} |\widehat{f}(x) + \widehat{f}(y)|^2 K(x, y) \mu(x) \\
&= \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \left| \sum_{a \in \Omega} P_x[a] f(a) + \sum_{b \in \Omega} P_y[b] f(b) \right|^2 K(x, y) \mu(x) \\
&= \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \left| \sum_{a,b \in \Omega} P_{x,y}[a, b] (f(a) + f(b)) \right|^2 K(x, y) \mu(x) \\
&\leq \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \sum_{a,b \in G} P_{x,y}[a, b] (f(a) + f(b))^2 K(x, y) \mu(x)
\end{aligned}$$

Using inequality (18), this gives

$$R_3 \leq \sum_{x,y \in \widehat{\Omega} \setminus \Omega} \sum_{a,b \in G} P_{x,y}[a, b] \sum_{\gamma \in \Gamma_{a,b}} G_{a,b}(\gamma) k[\gamma] \sum_{i=0}^{k(\gamma)-1} (f(v_{a,b,i+1}) + f(v_{a,b,i}))^2 K(x, y) \mu(x)$$

In particular, the coefficient of $(f(q) + f(r))^2$ in this upper bound is

$$(20) \quad [(f(q) + f(r))^2] R_3 \leq \sum_{\gamma_{a,b} \ni (q,r)} G_{a,b}[\gamma_{a,b}] t_{(a,b)}(\gamma_{a,b}) k[\gamma_{a,b}] \sum_{(x,y) \in \widehat{E}, x,y \notin G} P_{x,y}[a, b] K(x, y) \mu(x)$$

Combining inequalities (17), (19) and (20), the coefficient of $(f(q) + f(r))^2$ in $R_1 + 2R_2 + R_3$ is bounded by

$$\begin{aligned}
[(f(q) + f(r))^2] (R_1 + 2R_2 + R_3) &\leq \sum_{\gamma_{x,y} \ni (q,r)} G_{x,y}(\gamma) t_{(q,r)}(\gamma) k[\gamma] K(x, y) \mu(x) \\
&\quad + 2 \sum_{\gamma_{x,z} \ni (q,r)} t_{(q,r)}(\gamma) k[\gamma_{x,z}] G_{x,z}[\gamma_{x,z}] \sum_{y \notin G} P_y[z] K(x, y) \mu(x) \\
&\quad + \sum_{\gamma_{a,b} \ni (q,r)} G_{a,b}[\gamma_{a,b}] t_{(q,r)}(\gamma) k[\gamma_{a,b}] \sum_{(x,y) \in \widehat{E}, x,y \notin G} P_{x,y}[a, b] K(x, y) \mu(x)
\end{aligned}$$

On the other hand, the coefficient of $(f(q) + f(r))^2$ in $\mathcal{F}_Q(f, f)$ is at least $Q(q, r) \nu(q)$. Thus, setting

$$\begin{aligned}
\mathcal{A} = & \sup_{Q(q,r)>0} \frac{1}{Q(q,r)\nu(q)} \left(\sum_{\gamma_{x,y} \ni (q,r)} G_{x,y}(\gamma) t_{(q,r)}(\gamma) k[\gamma] K(x,y) \mu(x) \right. \\
& + 2 \sum_{\gamma_{x,z} \ni (q,r)} t_{(q,r)}(\gamma) k[\gamma_{x,z}] G_{x,z}[\gamma_{x,z}] \sum_{y \notin G} P_y[z] K(x,y) \mu(x) \\
& \left. + \sum_{\gamma_{a,b} \ni (q,r)} G_{a,b}[\gamma_{a,b}] t_{(q,r)}(\gamma) k[\gamma_{a,b}] \sum_{(x,y) \in \hat{E}, x,y \notin G} P_{x,y}[a,b] K(x,y) \mu(x) \right)
\end{aligned}$$

this implies

$$\mathcal{F}_K(\hat{f}, \hat{f}) \leq \mathcal{A} \mathcal{F}_Q(f, f)$$

which completes the proof. ■

8. COMPARISON FOR CHAINS ON CONTINUOUS STATE SPACES

In this section, we write down an analogue to Theorem 4 for Markov chains on continuous state spaces, based on Theorem 3.2 of [24]. It is necessary to develop some notation and definitions.

We consider two state spaces $S \subset \hat{S}$, with measurable sets \hat{F} and $F = \{A \cap S | A \in \hat{F}\}$. Then let $K(x, dy)$ and $Q(x, dy)$ be measurable kernels on (\hat{S}, \hat{F}) and (S, F) with stationary distributions μ and ν . Again, the goal will be to describe the mixing properties of Q in terms of the mixing properties of K , using spectral information. Although this setup is quite general, and much of the work goes through in greater generality, we will assume that S and \hat{S} are Lebesgue-measurable subsets of \mathbb{R}^n . We will write dx for a reference measure on \hat{S} , and we will also assume that S has nonzero measure under dx . In particular, if S is a submanifold of a manifold \hat{S} , we allow \hat{S} to have positive codimension in \mathbb{R}^n , but don't allow S to have positive codimension in \hat{S} . This zero-codimension assumption cannot be dropped easily; the Markov kernel on \hat{S} will generally give no information about kernels on subsets of measure 0.

Say that a Kernel P with stationary distribution π is reversible if $\pi(dx)P(x, dy) = \pi(dy)P(y, dx)$. Note that if P is reversible, it is a self-adjoint operator on $L^2(\pi)$, and so in particular has a real spectrum. Let $\lambda_0(P)$ be the infimum of the spectrum of P on the orthogonal complement of $\mathbf{1}$, and let $\lambda_1(P)$ be the supremum of this spectrum. As in the discrete case,

$$\lambda_1(P) = \inf \left\{ \frac{(f, (I - P)f)_\pi}{\|f\|_{L^2(\pi)}^2} : (f, \mathbf{1})_\pi = 0, f \neq 0 \right\}$$

Say that a kernel P is α -lazy if we can write $P(x, dy) = \alpha\delta_x(dy) + (1 - \alpha)\mu_x(dy)$, where δ_x is the measure concentrated at x and $\mu_x(dy)$ is any measure. If P is $\frac{1}{2}$ -lazy, then $\lambda_0(P) \geq 0$, and so can essentially be ignored. For $\epsilon > 0$, also define $\|\gamma_{x,y}\|_\epsilon = \sum_{(u,v) \in \gamma_{x,y}} (k_u(v)\rho(u))^{-2\epsilon}$. Unlike the discrete case, a bound on $\lambda_1(P)$ doesn't immediately give a bound on the total variation distance. Instead, we have only

$$\|\tau P^t - \pi\|_2 \leq \|\tau - \pi\|_2 |\lambda_1(P)|^n$$

It is now possible to set up the main comparison theorem. Assume that K and Q are $\frac{1}{2}$ -lazy, and furthermore that we can write $K(x, dy) = \frac{1}{2}\delta_x + k_x(y)dy$, $Q(x, dy) = \frac{1}{2}\delta_x + q_x(y)dy$, $\mu(dy) = \rho(y)dy$ and $\nu(dy) = \tau(y)dy$ for the reference measure dy on \hat{S} . As in the discrete theory, the first step is to define for all functions f on S their extensions \hat{f} to \hat{S} . Define for all $x \in \hat{S} \setminus S$ a measure $r_x(a)da$ on S , and set $\hat{f}(x) = \int_S f(a)r_x(a)da$. It is also necessary to define couplings $r_{x,y}(a,b)dadb$ of the measures $r_x(a)da$ and $r_y(b)db$.

Finally, it is necessary to define paths. This is slightly more complicated than the continuous situation. For fixed $x, y \in \hat{S}$ and kernel K , let a sequence $x = v_0, v_1, \dots, v_k = y$ be called a *path* from x to y if $k_{v_i}(v_{i+1}) > 0$ for all $0 \leq i < k$. Say that (x, y) require a path if $r_{ab}(x, y) > 0$ for some pair a, b with $\nu(a)k_a(b) > 0$, and denote by $\mathcal{P} \subset S^2$ the collection of pairs requiring a path. Then for fixed x, y , let Γ_{xy} be the collection of paths from x to y , and let $G : (x, y) \rightarrow \gamma_{xy} \in \Gamma_{xy}$ be a choice of a single element $\gamma_{xy} \in \Gamma_{xy}$ for each pair $(x, y) \in \mathcal{P}$. For the fixed $\gamma_{xy} \in \Gamma_{xy}$, let $|\gamma_{xy}|$ be the number of elements in γ , and let $\gamma_{xy}[i]$ be the i 'th element. Unlike the discrete case, some regularity assumptions are also needed.

Let $V = \{(x, y, i) : (x, y) \in \mathcal{P}, 1 \leq i \leq |\gamma_{xy}|\}$. Say that G satisfies the first regularity condition if the map $T(x, y, i) = (G(x, y)[i-1], G(x, y)[i], |G(x, y)|, i)$ from V to $S^2 \times \mathbb{N}^2$ is injective. Then, for all $b, i \in \mathbb{N}^2$ such that $(u, v, b, i) \in T(V)$ for some $u, v \in S^2$, let $W_{b,i} = \{(u, v) : (u, v, b, i) \in T(V)\}$. Assume that G satisfies the first regularity condition, and define the 1 to 1 map $H_{bi} : W_{b,i} \rightarrow S^2$ given by $H_{bi}(u, v) = (x, y)$ where $T(x, y, i) = (u, v, b, i)$. Say that G satisfies the second regularity condition if H_{bi} can be extended to a bijection of open sets with continuous partial derivatives a.e. with respect to Lebesgue measure. For the remainder of this paper, we will denote this extension by H_{bi} .

Assuming the two regularity conditions hold, define, for all $b, i \in \mathbb{N}^2$ such that $(u, v, b, i) \in T(V)$ for some $u, v \in S^2$, $J_{bi}(u, v)$ to be the Jacobian of the change of variables $H_{bi}(u, v) = (x, y)$. Note that these regularity conditions make the continuous comparison theorem substantially harder to use; they mean that a small change in an edge must result in only a small change of the path between its endpoints. Despite this, continuous versions of Theorems 5 and 6 are still easy.

Theorem 12 (Comparison for Chains on Continuous State Spaces). *Under the conditions described above, for all $\epsilon \in \mathbb{R}$,*

$$((I - K)\widehat{f}, \widehat{f})_\mu \leq \mathcal{A}_\epsilon((I - Q)f, f)_\nu$$

where

$$\begin{aligned} \mathcal{A}_\epsilon &= \text{esssup}_{(u,v) \in E} \{ (q_x(y)\tau(u))^{-(1-2\epsilon)} \\ &\quad \times \sum_{\gamma_{x,y} \ni (u,v)} \|\gamma_{x,y}\|_\epsilon k_x(y) \rho(x) |J_{xy}(u, v)| \} \\ &+ \text{esssup}_{(u,v) \in E} \{ (q_x(z)\tau(u))^{-(1-2\epsilon)} \\ &\quad \times \sum_{\gamma_{x,z} \ni (u,v)} \|\gamma_{x,z}\|_\epsilon \left(\int_{\widehat{S} \setminus S} r_y(z) k_x(y) dy \right) \rho(x) |J_{xz}(u, v)| \} \\ &+ \text{esssup}_{(u,v) \in E} \{ (q_a(b)\tau(u))^{-(1-2\epsilon)} \sum_{\gamma_{a,b} \ni (u,v)} \|\gamma_{a,b}\|_\epsilon \left(\int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} r_{xy}(a, b) \rho(x) k_x(y) dx dy \right) \\ &\quad \times \rho(a) |J_{ab}(u, v)| \} \end{aligned}$$

Proof. Start by writing

$$\begin{aligned} ((I - K)\widehat{f}, \widehat{f})_\mu &= \frac{1}{2} \int \int_{\widehat{S} \times \widehat{S}} (\widehat{f}(x) - \widehat{f}(y))^2 \rho(x) k_x(y) dx dy \\ &= \frac{1}{2} \int \int_{S \times S} (f(x) - f(y))^2 \rho(x) k_x(y) dx dy \\ &\quad + \int \int_{S \times \widehat{S} \setminus S} (f(x) - \widehat{f}(y))^2 \rho(x) k_x(y) dx dy \\ &\quad + \frac{1}{2} \int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} (\widehat{f}(x) - \widehat{f}(y))^2 \rho(x) k_x(y) dx dy \\ &\equiv \frac{1}{2} R_1 + R_2 + \frac{1}{2} R_3 \end{aligned}$$

The goal is to compare this to $((I - Q)f, f)_\nu = \frac{1}{2} \int \int_{S \times S} (f(x) - f(y))^2 \tau(x) q_x(y) dx dy$. R_1 is bounded exactly as in Theorem 3.2 of [23]:

$$\begin{aligned} \frac{1}{2} R_1 &= \frac{1}{2} \int \int_{S \times S} (f(x) - f(y))^2 \rho(x) k_x(y) dx dy \\ (21) \quad &\leq \text{esssup}_{(u,v) \in E} \{ (q_x(y)\tau(u))^{-(1-2\epsilon)} \\ &\quad \times \sum_{\gamma_{x,y} \ni (u,v)} \|\gamma_{x,y}\|_\epsilon k_x(y) \rho(x) |J_{xy}(u, v)| \} ((I - Q)f, f)_\nu \end{aligned}$$

Next, we bound R_2 by writing

$$\begin{aligned}
R_2 &= \int \int_{S \times \widehat{S} \setminus S} (f(x) - \widehat{f}(y))^2 \rho(x) k_x(y) dx dy \\
&= \int \int_{S \times \widehat{S} \setminus S} (f(x) - \int_{z \in S} r_y(z) f(z) dz)^2 \rho(x) k_x(y) dx dy \\
&= \int \int_{S \times \widehat{S} \setminus S} \left(\int_{z \in S} r_y(z) (f(x) - f(z)) dz \right)^2 \rho(x) k_x(y) dx dy \\
&\leq \int \int_{S \times \widehat{S} \setminus S} \int_{z \in S} (f(x) - f(z))^2 r_y(z) dz \rho(x) k_x(y) dx dy \\
&= \int \int_{S \times S} (f(x) - f(z))^2 \rho(x) \left(\int_{\widehat{S} \setminus S} r_y(z) k_x(y) dy \right) dx dz
\end{aligned}$$

This last term is bounded exactly as in Theorem 3.2 of [23], with $k_x(z)$ replaced by $(\int_{\widehat{S} \setminus S} r_y(z) k_x(y) dy)$:

$$\begin{aligned}
R_2 &\leq \text{essup}_{(u,v) \in E} \{ (q_x(z) \tau(u))^{-(1-2\epsilon)} \\
(22) \quad &\times \sum_{\gamma_{x,z} \ni (u,v)} ||\gamma_{x,z}||_\epsilon \left(\int_{\widehat{S} \setminus S} r_y(z) k_x(y) dy \right) \rho(x) |J_{xz}(u,v)| \} ((I-Q)f, f)_\nu
\end{aligned}$$

Finally, we bound R_3 by writing

$$\begin{aligned}
R_3 &= \int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} (\widehat{f}(x) - \widehat{f}(y))^2 \rho(x) k_x(y) dx dy \\
&= \int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} \left(\int_{a \in S} f(a) r_x(a) da - \int_{b \in S} f(b) r_y(b) db \right)^2 \rho(x) k_x(y) dx dy \\
&= \int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} \left(\int_{a,b \in S} (f(a) - f(b)) r_{x,y}(a,b) da db \right)^2 \rho(x) k_x(y) dx dy \\
&\leq \int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} \int_{a,b \in S} (f(a) - f(b))^2 r_{x,y}(a,b) da db \rho(x) k_x(y) dx dy \\
&= \int \int_{S \times S} (f(x) - f(y))^2 \left(\int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} r_{xy}(a,b) \rho(x) k_x(y) dx dy \right) da db
\end{aligned}$$

This last term is bounded exactly as in Theorem 3.2 of [23], with $\rho(a) k_a(b)$ replaced by $(\int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} r_{xy}(a,b) \rho(x) k_x(y) dx dy)$:

(23)

$$R_3 \leq \text{esssup}_{(u,v) \in E} \{ (q_a(b)\tau(u))^{-(1-2\epsilon)} \sum_{\gamma_{a,b} \ni (u,v)} \|\gamma_{a,b}\|_\epsilon \left(\int \int_{(\widehat{S} \setminus S) \times (\widehat{S} \setminus S)} r_{xy}(a,b) \rho(x) k_x(y) dx dy \right) \times \rho(a) |J_{ab}(u,v)| \} ((I-Q)f, f)_\nu$$

The theorem follows by combining inequalities (21), (22) and (23). ■

9. APPLICATIONS TO THE SPECTRAL PROFILE

In this section, we prove Theorem 9 using the techniques found in [11]. First, we recall the notation in that paper. For $S \subset \Omega$, let $c_0(S) = \{f : f \geq 0, \text{supp}(f) \subset S, f \neq \text{const}\}$. Then define, for kernel Q with stationary distribution ν ,

$$\lambda(S) = \inf_{f \in c_0(S)} \frac{\mathcal{E}_Q(f, f)}{V_\nu(f)}$$

And let $\nu_{\min} = \min_{\omega \in \Omega} \nu(\omega)$. Then define the *spectral profile* associated with Q by:

$$\Lambda(r) = \inf_{\nu_{\min} \leq \nu(S) \leq r} \lambda(S)$$

Define the spectral profile $\widehat{\Lambda}$ associated with K analogously. The main use of this definition in this context is through the following immediate consequence of Corollary 2.1 of [11]:

Theorem 13 (Spectral Profile Bound). *Fix $\epsilon > 0$ and let X_t be a $\frac{1}{2}$ -lazy, reversible chain with associated spectral profile Λ . Then for $t > \int_{4\nu_{\min}}^{4\epsilon^{-1}} \frac{2}{r\Lambda(r)} dr$,*

$$\|\mathcal{L}(X_t) - \nu\|_{TV} \leq \epsilon$$

We will use this bound with the following lemma:

Lemma 14 (Comparison for Spectral Profile). *Let M be a matrix with nonnegative entries such that $Mf \in \mathbb{R}^{\widehat{\Omega}}$ is an extension of f for all $f \in \mathbb{R}^\Omega$. Assume that*

$$\mathcal{E}_Q(f, f) \geq \mathcal{A} \mathcal{E}_K(Mf, Mf)$$

Furthermore, set $C_1 = \sup_{x \in \Omega} \frac{\nu(x)}{\mu(x)}$. Finally, for $S \subset \Omega$, define $\widehat{S} \subset \widehat{\Omega}$ to be the support of $M\mathbf{1}_S$ and $C_2 = \sup_{S \subset \Omega} \frac{\nu(\widehat{S})}{\mu(\widehat{S})}$. Then

$$\Lambda(r) \geq \frac{\mathcal{A}}{C_1} \widehat{\Lambda}(C_2 r)$$

Proof. For all $f \in c_0(S)$,

$$\begin{aligned}\lambda(S) &\geq \frac{\mathcal{E}_Q(f, f)}{V_\nu(f)} \\ &\geq \frac{\mathcal{A}\mathcal{E}_K(Mf, Mf)}{C_1 V_\mu(Mf)}\end{aligned}$$

But by assumption, the support of Mf is contained in \widehat{S} . Thus,

$$\lambda(S) \geq \frac{\mathcal{A}}{C_1} \widehat{\lambda}(\widehat{S})$$

The result follows immediately. ■

We will now use this lemma along with Theorems 13 and 5 to prove Theorem 9. The distributions and flow will be as in the proof of Theorem 5; it is easy to check that, in the notation of Lemma 14, $C_2 \leq 4$ in this example. Thus, the only missing ingredient is a bound on the spectral profile $\widehat{\Lambda}$ associated with simple random walk on the torus. By remark 6 following Theorem 1.2 of [6], the simple random walk on the torus has a property known as $(\frac{1}{2}, 2)$ moderate growth (see that paper for a definition of *moderate growth*). By Lemma 5.3 of [7], this walk satisfies what is known as a local Poincare inequality, with constant 8 (again, see that paper for a definition of *local Poincare inequality*). We don't use these two properties directly, but combining them with the inequality following equation 4.3 of [11], the spectral profile of the random walk on the torus satisfies the inequality

$$\widehat{\Lambda}(r) \geq \left(\frac{8}{27r} - 1 \right) \frac{1}{2n^2}$$

Thus, by Lemma 14 and the comments immediately following it, along with Theorem 5, the spectral profile of the walk on the torus with holes satisfies

$$\Lambda(r) \geq \frac{9}{4n^2} \left(\frac{2}{27r} - 1 \right)$$

Theorem 9 follows immediately from this bound and Theorem 13.

REFERENCES

- [1] Olena Blumberg. Permutations with interval restrictions. *PhD Thesis, Stanford University*, 2011.
- [2] Olena Blumberg. Cutoff for the transposition walk on permutations with one-sided restrictions. *Preprint*, 2012.
- [3] Persi Diaconis, Ronald Graham, and Susan Holmes. Statistical problems involving permutations with restricted positions. *IMS Lecture Notes Monogr. Ser.*, pages 195–222, 1999.
- [4] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walks on finite groups. *Annals of Probability*, 21(4):2131–2156, 1993.

- [5] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible markov chains. *Annals of Applied Probability*, 3(3):696–730, 1993.
- [6] Persi Diaconis and Laurent Saloff-Coste. Moderate growth and random walk on finite groups. *Geom. Func. Anal.*, 4(1):1–36, 1994.
- [7] Persi Diaconis and Laurent Saloff-Coste. Nash inequaltites for finite markov chains. *Journal of Theoretical Probability*, 9:459–510, 1996.
- [8] Persi Diaconis and Laurent Saloff-Coste. Walks on generating sets of abelian groups. *Probability Theory and Related Fields*, 105:393–421, 1996.
- [9] Persi Diaconis and Laurent Saloff-Coste. Walks on generating sets of groups. *Inventiones Math*, 134(2):250–301, 1998.
- [10] Martin Dyer, Leslie Goldberg, Mark Jerrum, and Russell Martin. Markov chain comparison. *Probability Surveys*, 3:89–111, 2006.
- [11] Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 2006.
- [12] Phil Hanlon. A random walk on the rook placements on a ferrers board. *Electron. J. Combin.*, 3(2):24, 1996.
- [13] Richard Holley and Dan Stroock. Logarithmic sobolev inequalities and stochastic ising models. *J. Stat. Phys.*, 46:1159–1194, 1987.
- [14] Roger Horn and Charles Johnson. *Matrix Analysis*. Cambridge Univ. Press, 1985.
- [15] John Jiang. Mixing time of markov chains on finite and compact lie groups. *PhD Thesis, Stanford University*, 2012.
- [16] Gady Kozma. On the precision of the spectral profile. *Alea*, 2007.
- [17] Tzong-Yow Lee and Horng-Tzer Yau. Logarithmic sobolev inequality for some models of random walks. *Annals of Probability*, 26(4):1855–1873, 1998.
- [18] David Levin, Yuval Peres, and Elizabeth Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, Rhode Island, 2009.
- [19] N. Metropolis, A.W. Rosenbluth, Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
- [20] Ravi Montenegro. Duality and evolving set bounds on mixing times. *Alea*, 2007.
- [21] Anastasia Raymer. Mixing time of the 15 puzzle. *PhD Thesis, University of California Davis*, 2011.
- [22] Laurent Saloff-Coste. Random walks on finite groups. *Probability on Discrete Structures*, pages 263–346, 2004.
- [23] Wai Kong Yuen. Applications of geometric bounds to the convergence rate of markov chains on rn. *Stoch. Proc. Appl.*, 87:1–23, 2000.
- [24] Wai Kong Yuen. Applications of geometric bounds to convergence rates of of Markov chains and Markov processes on rn. *PhD Thesis, University of Toronto*, 2001.

ICERM, BROWN UNIVERSITY, PROVIDENCE, RI

E-mail address: asmith3@math.stanford.edu